

## What from Toronto next week?

*This year's economic summit, arranged for next week, is unlikely to break with the bland self-reassurance that has become its tradition. But the seven governments need more imagination than usual.*

The heads of the governments of the seven major industrialized economies, with their economic ministers, will be meeting in Toronto next week further to coordinate their economic policies. They will also pay some attention, in what amounts to a sideshow, to a handful of scientific issues. The genteel convention that they should include these topics on their agenda goes back to the meeting at Versailles in 1981, when the then newly elected President François Mitterrand urged his fellow heads of state to put their faith in research and the new technology it would spawn — and mostly earned the reputation of being a dreamer. Things have changed a lot since then. Several proposals for the better coordination of research have been proposed at economic summits, and have been prosecuted with varying degrees of vigour. This year, the sideshow agenda will include a discussion of Japan's Human Frontiers Science Program (see *Nature* 33, 488; 1988). It will not, so far as is known, include the more radical proposal for shifting the basis of the modern industrial economy put forward by E.G. Nisbet on page 617 of this issue.

One obvious impediment is psychological. The industrialized economies of the West are this year still puzzled, but much relieved, to know how and why they have escaped the recession that seemed to threaten them in the wake of the great stock-market crash on 19 October last. Then it seemed inevitable that economic growth, especially quick last year would be brought to a halt and (horror!) even made negative. So people will be travelling to Toronto suffused with a warm glow of contentment. The worst (a quick slump) has not happened; and the greater the distance in time between now and 19 October, the smaller the chance there will be a recession or, at least, that it will be attributable to the stock-market crash.

### Lesson

Over-contentment, even on this narrow issue, will be misplaced, even if one lesson of the economic history of the past nine months is that national banks have learned a great deal from the earlier history of the origins of the great depression of the 1930s. Briefly, they understood, in 1987, that economies suddenly faced with the suspicion of impoverishment by the collapse of equity prices must be supplied with money if they are not to teeter into slump; in 1931, the central banks took the opposite course, with disastrous results. But next week's Toronto meeting had better rack its brains on the outstanding question of whether and when the inflation that is one of the expected consequences of last year's monetary laxity will make its appearance, and the subsidiary question of what the consequences might be. For the time being, finance ministers almost everywhere are in a cleft stick; fending off inflation requires higher interest rates, but higher interest rates are said to inhibit economic activity and so risk precipitating recession. The prospect is in any case further complicated by the two outstanding imbalances in the international system — the huge indebtedness of the developing world and the growing indebtedness of the United States. The communique that emerges from Toronto must deal with these issues squarely if it is to carry conviction.

The link between science and economics is relevant even on

this narrow front. There may be more to say about the puzzle of the recession that has not (so far) materialized than to congratulate the central bankers on their sagacity. The stock-market crash, after all, coincided with three dramatic but continuing trends in the pattern of national economies — the increased proportions of personal incomes being spent on services rather than on manufactured goods, the remarkable growth of the telecommunications business, especially internationally, and the equally remarkable insinuation of computer technology throughout modern industry. There is at least a possibility that the new machines, as distinct from the smokestack industries of the 1930s, make it possible for people threatened with impoverishment to create wealth from nothing more tangible than their own ingenuity. And if that influence was effective in 1987, what other technical trends may head off similar troubles in the future? Again before they congratulate each other too enthusiastically, the finance ministers might commission a study of these important but also interesting questions.

### Japan

They should also pay more attention to Japan's Human Frontiers Science Program than will be their natural inclination. The received wisdom, that the programme is too vague a prospectus of research that might be done, almost deliberately misses the point. Originally conceived by the Ministry of International Trade and Industry, but latterly more directly influenced by the research agencies in Tokyo, the programme sprang from the conviction that imaginative research and development projects at the interface between biology and the rest of science might be of great benefit all round, in fields as different as robotics and the treatment of disease, for example. Much of the confusion of the past two years arises because Japan has been trying to kill several birds with the same stone — declarations that the programme would make some restitution for Japan's early dependence on borrowed science and technology led many to think of the programme as a kind of pork-barrel, advocacy of particular schemes such as the sequencing of the human genome has been mistaken for a subtle attempt to avoid an international competition to get "there" first (wherever that may be). And so on. Japan deserves a better hearing at Toronto.

Nisbet's demand on attention at Toronto is a different case. Not that there can be much dispute about the logic of his argument. If, indeed, there are developing features of the global environment that threaten to put a crimp in plans for economic growth, they deserve consideration by the world's economic planners. But the trouble, for the time being, is that the question Nisbet asks is hypothetical, of the type "What should we do if . . . ?" The question is merely whether an economic summit like that arranged for next week is the place at which to start the ball rolling. Over the years, too many serious and strictly economic problems have been papered over with platitudes at meetings like this. But what the world needs is a substantial study of the ways in which the industrialized countries might economically adapt to a changing climate, or fend off change. The sooner that is done, the sooner is climatic change likely to appear on the economists' agenda. □



# New French government's windfall for research

- Mitterrand government backs research
- "Flagrant deficiencies" in funding

## Paris

FRENCH researchers, who have been feeling ill-treated following two years of government cutbacks in public sector spending, have been given a first taste of what to expect under President François Mitterrand's newly appointed socialist government. Minister-delegate for research Hubert Curien, last week announced that it is too late to cancel the "inadequate" 1988 research budget put together by his predecessor, Jacques Valade. Instead, an extra FF1,230 million (£117 million) will be handed out immediately to halt a net downward trend in research spending since 1986. This injection of cash is in addition to a FF350 million bonus released last week for higher education.

At a press conference on 8 June, Curien did not hide his disapproval of the previous government's research policy. As minister for research when the Fabius government lost power in 1986, Curien saw his long-term plans gradually dismantled over the following two years. Whereas the research budget had been set to increase by 4 per cent annually from December 1985, it has, when inflation is

taken into account, decreased by 6.6 per cent. "This slow-down", says Curien, "will make it impossible to reach the goal we had set of spending 3 per cent of Gross Domestic Product on research by 1990."

The extra cash will be used to underline priorities in the new government's research policy — jobs for researchers and opportunities for collaboration between industry and public-sector research establishments. These measures will be good news for young postdoctorates and doctoral students. Not only are postgraduate grants to be increased by more than one third to FF7,000 per month (affecting 4,630 students), but 150 new research posts are to be created this year. Of these, 85 have been earmarked for the national space research centre (CNES) so that young postdocs will be able to contribute to the long-term space plan approved by the European Space Agency last year. The remaining 65 jobs will be divided between the agricultural research institute (INRA), for biotechnology research, and the medical research institutes (INSERM), for work on cancer, immunology and epidemiology.

Curien has pledged to maintain a policy

of job creation from 1989 in order to offset a gradual ageing among the research community due to a recruitment boom in the 1960s which was not followed up consistently. Similarly, by increasing post-graduate grants, Curien hopes to save the "most brilliant" students from disappearing into industry rather than carrying out work for a PhD.

Of the FF90 million set aside for job creation and training, FF25 million is available for the rest of 1988 to enable about 200 foreign researchers to visit French laboratories. Curien has also cancelled the suspension of 150 engineer and technician posts called for by Valade (one out of every two vacancies) which had "threatened to paralyse several research teams" dependent on complicated instruments.

Curien's faith in the importance of industrial and technology research is well-known to the French science community. Since this faith also underpinned the previous government's policy of reducing public spending and increasing tax incentives for industrial research and development, many were curious to see how Curien's strategy would differ. The emergency allocations have given some strong clues.

While leaving Valade's FF500 million tax incentive scheme untouched, Curien has doubled the money available this year to the national research and technology fund (FRT) for 11 national priority programmes set up by the previous government (see *Nature* 329, 380; 1987). Under this scheme, awards are made following a call for tenders from private or public sectors. But, says Curien as there was not enough money in the kitty to cope with the demand, only one sixth of the proposals could be financed. Of the projects left on the shelf, those dealing with new ceramic fibres, multi-coated glass and metallurgy are now likely to be supported.

Referring to "flagrant deficiencies" in the money available to "certain organizations and ministries", Curien has promised that medical (especially molecular biology), agricultural and road-safety research will receive a slice of the extra FF725 million available this year. Similarly, ANVAR, the government agency that promotes collaboration between industry and state research institutes will get an extra FF130 million this year. When the conservative government took office in 1986, ANVAR's budget was cut from FF990 million to FF590 million.

Curien has also saved the Oceanographic Vessel of the Future (NCF), being built by IFREMER, the national institute for oceanographic research, from languishing in dry-dock. An extra FF80 million has been released to buy essential equipment, such as a multi-beam sonar probe, that could not be afforded.

Peter Coles

## French education at all levels a priority

### Paris

JUST a few days before the French people went to the polls for the third time in two months — this time to elect representatives for the national assembly — President François Mitterrand kept his election promise of more money for education. The new education and research minister, Lionel Jospin, announced on 1 June after a cabinet meeting that FF1,200 million (£120 million) "emergency funds" are immediately to be made available to reduce "black spots, serious gaps and delays" within secondary and higher education.

About 30 per cent (FF350 million) of the new money will go towards higher education, especially to renovate ageing university buildings (FF100 million) and halls of residence (FF50 million). An extra FF50 million will also be given to libraries for book purchase and documentation.

The statute that allows governments to vote emergency funds cannot be used to create new posts — as education unions want — but 600 higher education administrator jobs cancelled by the previous government will be restored. Jospin has

also reversed a move by his predecessor, Jacques Valade, to reintroduce the two-tier doctorate abandoned in 1984 in favour of the PhD (see *Nature* 332, 388; 1988).

For the longer term, Jospin has confirmed that the new government's 1989 budget will include an extra FF4,000 million for education, to be repeated each year until 1992 in order to reach the FF40,000 million target promised by Mitterrand in his presidential campaign.

Jospin has also decided to allow his minister delegate for research, Hubert Curien, to take an active part in the French space programme — which he helped shape under the previous socialist government (see *Nature* 333, 197; 1988). Responsibility for space in Mitterrand's new government was taken away from the research ministry and given to Paul Quilès, minister for posts and telecommunications, Curien will also be jointly responsible, with industry minister, Roger Faroux, for the ultra-modern science and technology museum, La Villette, as well as for the government's technology innovation body, ANVAR.

Peter Coles



# Mandatory AIDS tests on basis 'of slight suspicion' in Bavaria

## Munich

DEEPLY disturbed by the civil rights implications of a Bavarian court decision announced on 31 May, three West German AIDS foundations are considering an appeal to the nation's highest court.

The Bavarian court, the Bayerischer Verwaltungsgericht, declared in an oral statement that a person can be forced by local health officials to be tested for antibodies to the human immunodeficiency virus (HIV) on the basis of even a "slight suspicion" of being infected. HIV is the virus that causes AIDS. Previously, a test could be required only if police proved to health officials that the person had been a member of a risk group such as intravenous drug users.

The case involves a man from the Rosenheim area who was an intravenous drug user until 1983. He had been tested for HIV with negative results in February 1985 after undergoing therapy. Another test was called for by police in June 1987, after the man had been found in possession of illegal but non-intravenous drugs such as hashish. The man then sought protection from the courts. A lower court upheld his refusal, but this ruling was overturned by the 31 May decision.

The foundations — the Deutsche AIDS-Stiftung (DAS), the Bayerische AIDS-Stiftung and the Nationale AIDS-Stiftung — view the higher court's ruling with mixed feelings. The court did not base its decision on the controversial "AIDS catalogue" of the Bavarian government, but rather solely on the federal epidemic law (Bundesseuchengesetz). This undermines the validity of the Bavarian measures, which have until now been condemned by the West German government and the other 10 *Länder* (states) (see *Nature* 327, 264; 1987). But the court's interpretation of the federal epidemic law might cause the other *Länder* to re-evaluate their policies and allow mandatory tests. Further clarification is expected when the court releases a written brief on the case.

The court justified its decision orally based on its judgement that the public's right to protection from AIDS outweighs the infected person's right not to know about his infection. The statement did not touch on the civil rights issues involved.

The ruling will not necessarily affect the way in which other cities in Bavaria interpret the law. In Munich, Senior City Administrator Hans-Peter Uhl stands by a "humane application" of the law which requires that the police give evidence of intravenous drug use or other risky behaviour before he will call for a test. Without such a humane approach, Uhl

said, there is a risk of returning to the "dark Middle Ages" when diseased people were "driven beyond the city wall." Uhl said that prostitutes and foreigners in Munich have been very cooperative about being tested in the year since the law took effect.

A legal challenge to the recent ruling may be difficult or even undesirable, said Ulrich Heide, managing director of the DAS in Cologne. The case cannot be brought to the highest German court, the Federal Constitutional Court (Bundesverfassungsgericht), by anyone but the

original defendant or the Bundestag (federal parliament). But by the time the Rosenheim man can make a legal challenge, the test may already have been carried out. Heide said that the DAS would not push the man to make a challenge, though it would probably help pay the legal costs in such a case.

A challenge might be undesirable because of the perceived conservative bias of the Federal Constitutional Court. Even a discussion of the matter by the Bundestag might be harmful in Heide's opinion. The best outcome, he said, would be a return to the "*status quo ante* of broad political consensus for education and information". But this is unlikely to be the last case of its kind. **Steven Dickman**

# Sober side up at giant AIDS meeting held in Stockholm

## Stockholm

ALMOST 100,000 cases of AIDS have now been reported to the World Health Organization (WHO) and probably an equal number are unreported, said Jonathan Mann, director of WHO's Global Programme on AIDS at the opening of the IV International Conference on AIDS in Stockholm last Sunday. He predicts that there will be at least a million new cases over the next five years.

Of these, some 300,000 are likely to be in the United States according to the latest predictions of the US Centers for Disease Control (CDC), although the confidence limits on these predictions are far from high. Even now, said CDC director James Curran, one new US case of AIDS is being reported every 14 minutes.

For Europe, Jean-Baptiste Brunet, of the Claude Bernard Hospital in Paris, predicted a cumulative total of 56,400 AIDS cases by the end of 1989, with a large increase in the proportion of cases associated with intravenous drug use, particularly in southern Europe.

No predictions are possible for Africa for a variety of reasons, one of which is that the availability of adequate figures on the proportion of people already infected with the human immunodeficiency virus (HIV) is patchy. At least in Kinshasa, capital of Zaïre, it is certain that between about 3 and 7 per cent of blood donors, pregnant women and factory or bank workers are infected. The brighter side of these figures is that they have remained constant for up to four years, said Bosenge N'Galy from Zaïre's Department of Public Health.

As if the epidemiologist's figures were not sobering enough, Sweden's Prime Minister Ingvar Carlsson had opened the conference by warning the delegates of winning the battle against AIDS in the laboratory but losing it in the streets. And

Robert Gallo of the National Cancer Institute in Bethesda said that laboratory research was now in a period of steady incremental progress rather than the dramatic advances of the first few years.

Nevertheless, both Gallo and Luc Montagnier from the Pasteur Institute were able to provide delegates with a few new facts to whet their appetites. Montagnier referred to emerging evidence that rhesus macaque monkeys infected with HIV-2, the predominant HIV in West Africa, develop AIDS-like symptoms. If this is confirmed, AIDS research may at last have an animal model. He also predicted that serum antibodies to the product of the HIV *nef* gene (to use the new nomenclature — see *Nature* 333, 504; 1988) would become an earlier indication of HIV infection than other antibodies.

Gallo drew attention to accumulating evidence that human herpes virus 6, recently discovered in his laboratory, is a cofactor in the progression from HIV infection to AIDS, and to laboratory data that suggest that Kaposi's sarcoma in AIDS patients is not the result of a second infectious agent. Instead it seems to be caused by a novel growth factor released by HIV-infected cells.

On the prospect of vaccines, Gallo was less bullish than in previous years, although still "confident that the problem will be solved". And Lars Olof Kallings, chairman of the conference, expressed disappointment that so few new compounds with activity against HIV were due to be reported.

The conference was attended by 7,000-plus delegates with 9 simultaneous sessions of workshops and round tables and over 3,000 posters. Delegate registration was greater than last year, with about 2,500 from North America and even more than that from Europe. **Peter Newmark**



# Uprising on West Bank places strains on academics

## Jerusalem

THE *intifada*, or uprising, in the Israeli-occupied West Bank and Gaza Strip has forced both Palestinian and Israeli academics to make difficult choices. Palestinians have expressed solidarity with the new leadership. For their Jewish Israeli counterparts, the six-month-old uprising has exacerbated the split between left and right.

Calls for commercial strikes, stone-throwing *mêlées* and defiance of Israeli tear gas and rubber bullets have brought life in the territories to a virtual standstill. The West Bank universities have been shut by the Israelis since February.

The students have caused "a breakdown of the social structure" according to Dr Baeb Erakat, lecturer in political science at A-Najah University in Nablus.

But, "the social transformation has brought [us all] closer together", he said, noting that it is the "generation of occupation, of defiance, which is leading the uprising". In fact, 77 per cent of the Palestinian population in the occupied territories is under 28 years old.

"The general tendency [among Palestinian academics] now is to go back to the people, to seek out their roots", says Dr Azmy Bishara of the philosophy department of the West Bank's Bir Zeit University.

So far, only one general proposal has been announced by a group of Palestinians that includes two academics. The 14-point plan, unveiled three months ago, calls for an end to the occupation and condemns the mass arrests, deportations and other "illegal steps" taken by the Israeli army, which "contravene the Geneva Convention".

Gabi Baranki, professor of chemistry, acting president of Bir Zeit and one of the plan's signatories, says, "I don't believe peace is something to keep on talking about. Concrete steps must be taken."

But the Jewish academic community is further from a common political view than ever. A few weeks after the disturbances began, 600 academics banded together as the 'Inter-University Movement for a Political Settlement', and signed a petition calling on the Israeli government to negotiate with the Palestinians.

The group is still debating the extent to which it is prepared to recognize the Palestine Liberation Organization (PLO). It did, however, recently issue a statement requesting that Israel end its policy of not talking to the PLO.

Contact with the PLO, however, is anathema to a right-wing group that emerged in reaction to the first petition. 'Professors for a Strong Stand' came out in

support of the army's tactics, which have left many Palestinians dead. Their petition boasted 700 signatures although they have been accused of turning to hospital physicians and other 'non-academics' to boost their numbers.

"We got irritated by those who came out in the name of the academic community, condemning the army for its actions in the territories", said Professor Hefzibah Eyal-Giladi, of the Hebrew University's biology department, who believes that Israel's reaction to the *intifada* "was too little, too weak".

One point made by Eyal Giladi and shared by many other Jewish Israelis is that there is no one to negotiate with. "The PLO is a tricky business; what it says one day is not the same as another day", says Leo Sachs, head of the Weizmann Institute's biology department. But Sachs also says that there must be a *modus vivendi* between the two communities, and that although the search for a settlement is bound to take time, it would be a "good idea" if the academic community

## Apology

IN a News article entitled "Japanese AIDS scandal over trials and marketing of coagulants" (*Nature* 331, 552; 1988), we stated that Professor Takeshi Abe of Teikyo University had admitted delaying approval of clinical trials and marketing of heat-treated blood coagulants for haemophiliacs so that a Japanese blood-product manufacturer could catch up with foreign competition.

We now understand this statement to be incorrect. Professor Abe helped to coordinate clinical trials of heat-treated coagulants and has explained that the advice he gave to the companies involved was intended to speed licensing. He further explained that the need for coordination, intended to ensure that licences would be granted without delay by the ministry's new drugs committee, may have given rise to the impression that some activities were being held back.

We also reported the *Mainichi* newspaper as saying that Professor Abe solicited donations from the companies involved in the clinical trials during the time they were trying to win approval for trials. We fully understand that all donations given to Professor Abe had no connection with the execution of the trials and were used to help realize Professor Abe's dream of establishing a foundation to aid haemophiliacs.

*Nature* regrets these errors and apologizes to Professor Abe for them. □

were to develop a workable plan.

The Palestinian academics, for their part, claim the right of their people to choose their own leaders, and reject Israeli moves to do this for them. As Erakat points out, a 1986 poll shows that 93 per cent of the Palestinian people support the PLO as their sole legitimate representative, so the imbroglio is not likely to end soon.

Lisa Perlman

## Vote-catching oil debate in California

### Berkeley

OPPONENTS of the plans of the US Department of the Interior for oil exploration off the northern California coast have gained a temporary victory, thanks in part to the forthcoming presidential election. The department has agreed to postpone the plan, but most agree that the delay is politically motivated, and may end after the election.

The offshore leasing plan has come under harsh criticism not only by Californians, but also by the Environmental Protection Agency and the US Fish and Wildlife Service, and most recently by presidential candidates courting the Californian vote.

Democratic candidate Michael Dukakis has promised to cancel the sale of leases for oil exploration on 1.1 million acres off the Mendocino coast. Republican candidate George Bush, an ex-Texas oilman who supports offshore drilling, recently said he would reserve his stand on the California plan until more is known about potential environmental impacts. Opponents of the drilling plan see Bush's change of heart as a political move in response to reports that California voters favour Dukakis. Most expect that Bush would proceed with the oil exploration if elected.

The day after Bush's announcement, and apparently in response to Republican pressure, Interior Secretary Donald Hodel called for an indefinite suspension of the lease sale, originally scheduled for February 1989. Hodel cancelled publication of a final environmental impact report that was due out in August.

There has been much criticism of the preliminary environmental impact report, published last December. An Environmental Protection Agency request for further habitat protection (see *Nature* 332, 673; 1988) was followed by even harsher criticism by the US Fish and Wildlife Service, an agency of the Interior Department.

Last week, the interior subcommittee of the House Appropriations Committee voted, without discussion, to delay any drilling in the area until October 1989. The full committee is expected to support the decision, given the current political volatility of the issue.

Marcia Barinaga

# UK leukaemia study implicates nuclear reprocessing plants

## London

A NUCLEAR reprocessing complex at Dounreay, in Scotland, is the most likely cause of an increased incidence of leukaemia among young people living in the vicinity, a government-commissioned study has concluded. The independent Committee on Medical Aspects of Radiation in the Environment, chaired by Professor Martin Bobrow, head of paediatric research at Guy's Medical School in London, found that six cases of childhood leukaemia were registered in the area

between 1979 and 1984, when only one would have been expected. The precise reason for the anomaly remains elusive, however, and the committee was unable to attribute the increased leukaemia incidence to the low levels of radioactive discharge from the site.

Increased incidences of leukaemia among children have also been found around Britain's only other nuclear reprocessing plant, at Sellafield, in Cumbria. The committee's report concludes: "The evidence of a raised incidence of leukaemia

near Dounreay, taken in conjunction with that relating to the area around Sellafield, tends to support the hypothesis that some feature of the nuclear plants that we have examined leads to an increased risk of leukaemia in young people living in the vicinity of those plants."

The Dounreay site, acquired by the UK Atomic Energy Authority (UKAEA) in 1955, has operated three nuclear reactors and their associated fuel fabrication and reprocessing plants and is Britain's centre for fast reactor development.

The study examined leukaemia incidence between 1968 and 1984 in an area within 25 km of the nuclear complex. During that time six cases were registered among people aged 0-24, twice the number expected on the basis of national rates, but a figure that was nevertheless statistically insignificant. However, five of the cases occurred within a distance of 12.5 km from the site, a threefold excess. Most strikingly, all six cases within the full 25-km zone were registered in the last six years of the study period, a 12-fold increase on national rates and an observation the committee describes as "remarkable".

On the question of possible causes, the committee says that on present estimates, risk to the fetus and infants arising from discharge at the Dounreay site would not account for the increased incidence of leukaemia. It notes, however, that data concerning leukaemia risk following low-level radiation exposure *in utero* and in infancy are "limited". Furthermore, limited knowledge about the physical and chemical forms of radionuclides concerned, their distribution in the body and the relevant target tissue for leukaemogenesis, particularly for children, means that it remains a possibility that some feature of the discharges could generate much higher doses to the target organs than present estimates would suggest. One common feature of Dounreay and Sellafield is that actinides, which emit radiation with a high linear-energy transfer, are discharged from both sites in a much higher proportion than from nuclear power generating plants.

The committee has recommended investigation into a number of alternative possibilities, including unrecognized pathways of radionuclide transfer in the environment, and possible pre-conception effects of low-level radiation.

The UKAEA commended the report for its "high quality", but predictably emphasized that the study was not conclusive. The authority's chairman, Mr John Collier, pledged support for future studies but was keen to point out that "leukaemia clusters were identified long before the nuclear age and exist in countries with no nuclear plants, like New Zealand".

Simon Hadlington

# Problems over TPA patent for Genentech in Japan

## Tokyo

GENENTECH'S worldwide battle to defend its patents covering the clot-dissolving glycoprotein called tissue plasminogen activator (TPA) suffered another setback at the end of last month, when the company failed to win a temporary injunction to halt clinical trials by a Japanese competitor, Toyobo.

But part of Genentech's problems in Japan seems to be a judge who does not understand genetic engineering. Meanwhile, Toyobo is nearing completion of its clinical trials. The patent dispute is the first in Japan involving recombinant DNA technology; it is a revealing illustration of the differences between US and Japanese patent law.

Genentech's patent was announced by the Japanese patent office in April last year, when several companies immediately filed objections. Under Japanese law, the patent will become effective only if and when these claims are cleared — a process that is expected to take about another year. Although, in the United States, patents remain confidential until granted, Japanese patent law provides opportunities for public review and opposition before patents are registered.

As in the case of the fight against Genentech's patent in the British courts by the UK company Wellcome, Genentech's competitors are objecting to the very broad nature of the US company's Japanese patent. In Japan, patents are typically very narrowly defined, and this in part accounts for the huge numbers of patents filed by Japanese companies.

Meanwhile, several Japanese companies are racing to put recombinant TPA on the market as a drug for heart attacks. Mitsubishi Chemical and Kyowa Hakko have licensing agreements with Genentech, while Sumitomo Pharmaceutical has an agreement with Wellcome.

Toyobo, which has licensed technology from the US company Integrated Genetics, is thought to be Genentech's closest rival in Japan. In August last year, Genentech filed a suit in the Osaka local district court alleging patent infringement by Toyobo, and demanding a ban on Toyobo's production, marketing and sales promotion of recombinant TPA as well as on the conduct of clinical trials.

Japanese trials of such patent disputes are notoriously prolonged; Genentech's case is expected to take at least two or three years. That is why, in February, Genentech also filed for a temporary injunction to halt Toyobo's clinical trials, pending a decision in its case. Although applications for such injunctions are usually decided fairly rapidly, the judge ruled in April that the injunction should be discussed alongside the case of patent infringement.

Now a further delay has been imposed. A new judge has been assigned to the case and, despite a mound of documents submitted by Genentech and Toyobo, said at the latest hearing on 31 May (according to a report in *Nikkei Biotechnology*) that "The court does not yet understand the real point of contention. The detailed documents . . . are difficult to read. If possible the plaintiff and defendant should point out the problematical points". The next hearing is on 28 July.

Genentech brought in Nobel prize-winner Paul Berg to help explain its case in the April hearing. The company has also approached some Japanese academics for help, but they are generally reluctant to become involved in court hearings.

Meanwhile, Toyobo is in phase two of the clinical trials and may be able to apply to the Ministry of Health and Welfare for marketing approval before a decision on the temporary injunction is reached.

David Swinbanks



## Shoreham nuclear power plant shuts, Seabrook stumbles

*Boston*

Shoreham and Seabrook, the never-licensed nuclear power plants in New York and New Hampshire, have finally sunk under the weight of a decade's political and financial opposition.

Shoreham's death became official two weeks ago, when the Long Island utility LILCO formally agreed to sell the unopened \$5,000-million plant to New York state for \$1. As part of the agreement, the state will oversee the closure and, ultimately, the dismantling of the plant.

LILCO will, as a result, take a loss on paper of \$2,500 million, but will be allowed to pass some of the rest of the cost onto its consumers. LILCO's electricity rates, already the second highest in the United States, will also be raised by 5 per cent a year. Meanwhile, investors in the plant have already suffered much of the remaining financial loss — the price of LILCO's stock has plummeted during the plant's demise.

State officials have not yet decided what to do with the plant they have acquired for \$1. The state says plainly that it intends to dismantle the reactor, but how and when is far from clear. At present, there is no place in the United States at which to dispose of the fuel rods (radioactive because of low-level operation) and reactor parts; there will probably not be a licensed disposal site until next century.

The proposed high-level nuclear waste

repository, originally planned to open in 1996, has yet to be started. Indeed, it is not yet clear whether the proposed site in Yucca Mountain, Nevada, will meet the Energy Department's geological and environmental criteria.

Another possibility for New York state is to mothball the plant in such a way that it could be reopened at a future date if warranted. So far, no nuclear plant in the United States has been dealt with in this way, and observers say that the obstacles to such a course are high. For one thing, Shoreham would need to retain a licence, which itself costs close to \$1 million a year, while the cost of the skeleton staff that would be necessary to maintain the reactor's operating potential would be higher still.

Although its ultimate fate has yet to be decided, the Shoreham plant is now certain to be shut down. But New Hampshire's Seabrook plant is still on its deathbed. The latest development is that one of the plant's largest remaining investors announced its intention to pull out, sun-dering the fragile investment coalition on which the plant depends.

The governor of Massachusetts, Michael Dukakis, who is foremost in the race for the Democratic nomination for the presidential election in November, has consistently opposed Seabrook and, if elected president, would most certainly shut the plant down.

**Seth Shulman**

## Agreement on minerals

*Washington*

AN agreement for regulating the development of Antarctica's mineral resources was reached last week between 33 countries that have interests in the continent. No immediate plans have been laid to exploit Antarctica's mineral and petroleum reserves.

The agreement replaces an informal moratorium on mineral development that has been observed by all nations for the past eight years. It outlines a framework for controlling every aspect of development, from prospecting and exploration activities to mineral extraction and processing, and puts priority on protecting the Antarctic for research purposes.

The extent of Antarctica's natural mineral resources has not yet been determined, but because of its size, it is suspected to contain commercially valuable minerals and petroleum. The high cost of mining in the Antarctic environment will make development unlikely until sources elsewhere are depleted.

The agreement is the latest under the Antarctic treaty system, begun in 1959 by the 12 countries that first sponsored research expeditions to Antarctica. The initial treaty was set up to preserve Antarctica for pure research: it bans military and nuclear activities, and guarantees freedom to conduct scientific research to all countries that respect these bans, as long as they share their data. Since 1959, eight more countries have initiated research programmes and become parties to the original treaty. Eighteen additional countries have indicated their approval, but do not have a vote in decisions under the treaty system because they have not conducted Antarctic research. In 1980 a second treaty governing the harvesting of fish and other wildlife was signed.

Negotiations concerning the new agreement on mineral resources stretched over six years, and were complicated by territorial claims made by Argentina, Chile, Australia, Britain, France, Norway and New Zealand — seven of the original 12 signatory nations to the Antarctic treaty system. These "claimant" states will be balanced against non-claimants, including the United States and the Soviet Union, on the review commission established by the agreement to regulate development. The new agreement will attain treaty status once it is signed and ratified by 16 of the 20 voting parties in the Antarctic treaty system, expected by the end of the year.

Many environmental groups, including the Antarctic and Southern Ocean Coalition, a consortium of 176 organizations, oppose the agreement.

**Carol Ezzell**

## Rain-making experiment in India

*New Delhi*

THE Indian Department of Science and Technology (DST) has defended its financial support of an extraordinary rain-making experiment that was supposed to invoke the rain god through a 3,000-year-old religious ritual described in Vedas, the holy scriptures of the Hindus.

The ritual, called yagna, involves chanting of vedic hymns in praise of the god Varuna before a sacred fire kept burning with twigs of herbs and a large quantity of clarified butter (ghee).

The officially sponsored yagna took place last week at Mathura, about 100 km from New Delhi, and was performed by an 86-year-old vedic scholar who claims that nine of the twelve yagnas he has performed in the past eight years brought rain.

Sharma chanted verses from Sama Veda before a fire kept burning six hours a day for a week with 100 kg of sandalwood and 15 selected herbs, and an equal quantity of ghee. As loudspeakers blared the vedic chants, scientists of the Indian Meteorological Department (IMD) using scientific

equipment monitored changes in the atmosphere that Sharma claimed his exercise would bring about. The scientists also collected samples of aerosols from the fire that supposedly had the power to attract clouds and induce rain.

A spokesman for DST said the yagna was part of IMD's continuing programme of "cloud-seeding using ground-based generators". He said Sharma's proposal was funded only after approval by a committee of physicists and meteorologists as it presented an opportunity scientifically to evaluate the validity of the ancient Hindu rite for creating artificial rain.

The yagna fire in Mathura failed to bring even a drop of rain but has landed DST in the fire. The experiment is widely interpreted as official sanction to superstition at a time when Prime Minister Rajiv Gandhi is exhorting scientists to take India into the twenty-first century. The DST, which tried to keep the yagna a secret from the press, says it was only following "the scientific tradition of being objective and open-minded".

**K.S. Jayaraman**



# SDI an imperfect shield says congressional agency

## Washington

THE Reagan administration's plans for a ballistic missile defence (BMD) have once again come under attack from the congressional Office of Technology Assessment (OTA), which released a critical report last week.

The report says that relatively simple countermeasures by the Soviet Union could foil a large-scale ballistic missile defence system, and that neither the Pentagon nor its contractors have "adequately addressed" the prospect that such a system would have to operate in the face of the "mutual occupation of space by weapons of comparable capability".

The complexity of the Strategic Defense Initiative (SDI) is its major weakness, the OTA report claims, and the system might never be shown to function properly because of the inability to test it adequately. In some of its strongest wording, the report claims that there would be "a significant probability . . . that the first (and presumably only) time the BMD system was used in a real war it would suffer a catastrophic failure".

A summary of the current report was leaked to the press in April (see *Nature* 332, 767; 1988), but its full findings have only now been made public after two years of study and more than eight months of haggling with the Pentagon over whether parts of it could be declassified.

Because of continued differences between OTA and the Pentagon, the three final chapters — which deal explicitly with potential Soviet countermeasures — were omitted in the unclassified version, much to OTA's displeasure. Of the negotiations with the Pentagon, the report says, "OTA found the wheels of bureaucracy to run very slowly — when they turned at all".

The report's several findings include the following:

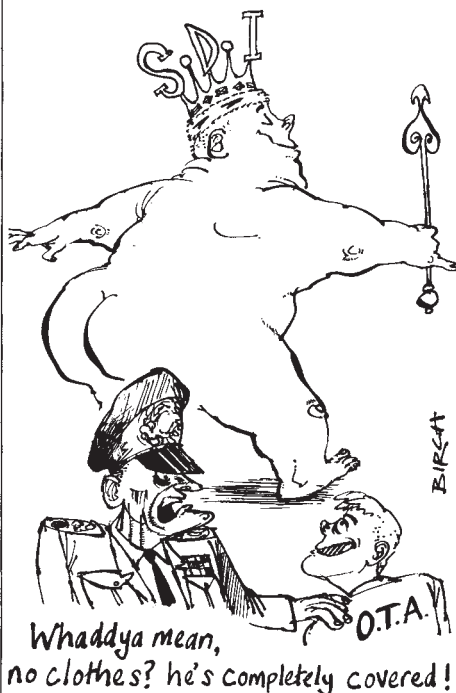
- Given optimistic estimates, such as "extraordinarily fast rates of research, development, and production", a limited BMD system might be technically deployable by the year 2000.
- The decision to deploy a limited BMD system which could destroy "anywhere from a few up to a modest fraction" of attacking Soviet missiles, would face the threat of quick obsolescence by Soviet countermeasures and would (at least to some degree) commit the United States to a system whose ultimate feasibility remains uncertain.
- The question of how well the system's software might work in a real battle situation would "always be irresolvable" which would clearly affect the confidence that the United States military could reasonably

place in the system's effectiveness when under fire.

The Pentagon, in a three-page response to the report, criticized OTA's findings as "unduly pessimistic". In particular, the Pentagon cited the report's conclusions about the potential for software problems as reflecting "opinion rather than real analysis".

A key premise of OTA's argument is that large-scale software systems, such as the long-distance telephone system, become dependable only after extensive use and modification. In contrast, such extensive testing would not be possible for a system which could only be used only in a nuclear war.

Despite the Pentagon's criticism, the OTA report is widely seen as one of the most thorough assessments of SDI so far made. Although most of its findings are



not new, the report has been praised for the diversity of its advisory panel, which included representatives from defence contractors, academic institutions and government, and has been lent further credence by the extensive access that the authors had to classified material during the preparation of the report.

Stephen Meyer, professor at the Massachusetts Institute of Technology and a member of the report's advisory panel says, "this study should play a particularly important role as a primer on the issue for the next Congress and Administration".

Seth Shulman

# Engineering lab next on UK privatization agenda

## London

THE British government's recently declared policy of shifting the financial burden of 'near market' research away from the public purse has been dramatically demonstrated with the announcement last week that the National Engineering Laboratory in East Kilbride, Scotland, is up for sale. The laboratory is one of three research establishments run by the Department of Trade and Industry (DTI). Bids for the laboratory, whose value and potential profitability the government has apparently not yet assessed, have been invited, with a deadline of 22 July. The suddenness of the announcement, by Trade Secretary Lord Young, and its implications for future employment prospects, provoked angry criticisms from representatives of the laboratory's workforce of 620.

The move has also sparked a political row, with allegations of impropriety in the way the sale is being handled. Earlier this year, a small private research firm, British Hydromechanics Research Association, was given permission by the DTI for a limited study of the laboratory's investment potential. The civil service union representing the laboratory's workforce and the local member of parliament want to know why such permission was granted in advance of the announcement of the sale.

The 40-year-old laboratory is involved in a wide range of advanced engineering research, including bioreactors, offshore engineering and wind energy. Of its annual budget of £21 million, £16 million comes direct from the government.

The DTI says that the laboratory is a prime candidate for privatization given that around three-quarters of its research is of direct relevance to industry.

The other three DTI research establishments are the Laboratory of the Government Chemist, which provides consultancy advice and studies based on chemistry, with special expertise in analytical chemistry, mainly for the public sector, and monitors the DTI's biotechnology activities; the Warren Spring Laboratory, which undertakes applied research and provides research services for industry, government and local authorities on a wide variety of process industry and environmental problems, including air and oil pollution, measurement and abatement; and the National Physical Laboratory, which provides measurement standards and calibration services.

The government wants these laboratories to remain in the public sector, but to limit the amount of work they do for private industry. Simon Hadlington

# Disputes delay development of leprosy vaccine for India

New Delhi

PROLONGED wrangling among Indian scientists is hampering efforts to produce a leprosy vaccine for the subcontinent. Four potential vaccines exist in varying states of development, but conflicting opinions, mutual distrust and competition between different groups of researchers is preventing the health ministry from deciding which to choose. The ministry has not yet given the go-ahead to what would be the world's largest trial of an armadillo-derived vaccine developed under the Immunology of Leprosy programme of the World Health Organisation (WHO). The vaccine is already under trial in Venezuela and Malawi, and it had been hoped to conduct a larger trial at the Jalma Institute of Leprosy in Agra, run by the Indian Council of Medical Research (ICMR).

The trial is being opposed by the Bombay-based Cancer Research Institute (CRI), which claims to have developed an equally effective and cheaper vaccine, derived from cultivable mycobacteria.

## Leishmaniasis resurgent in India

New Delhi

LEISHMANIASIS, an insect-borne disease, has staged a comeback years after it was wiped out from India. In the past 12 months, some 22,000 cases and 160 deaths have been reported from the states of West Bengal, Bihar and Uttar Pradesh. There have also been cases in Tamil Nadu and Delhi.

Leishmaniasis is caused by a protozoan, leishmania, passed on to humans by sandflies. Sandflies are extremely sensitive to DDT, and had virtually been eradicated as a side-effect of malaria control in the 1950s. But after 1965, insecticide spraying was stopped under the government's modified anti-malaria strategy, resulting in a build-up of the sandfly population and a resurgence of the disease. In 1977, leishmaniasis killed 229 people in Bihar and since then has spread to adjoining states.

According to the Health Ministry, leishmaniasis is unlikely to become an epidemic if preventive measures are continued. A suggestion by an expert committee for a national control programme has been rejected. Instead, affected states have been asked to re-start door-to-door DDT spraying. India is producing sodium antimony gluconate, the conventional drug for treating victims, and is importing pentamidine isothionate for the 10 per cent of cases that do not respond to antimony treatment.

K.S. Jayaraman

Opponents of the WHO scheme say that India should be encouraged to develop an indigenous vaccine, rather than depend on an imported one. The CRI vaccine is in the third phase of clinical trial in a leprosy-endemic district of Maharashtra. Some 10,000 people have been vaccinated, with a further 40,000 to be brought under the trial by 1991. CRI director Dr Mahadev G. Deo questions the need for an expensive trial of the WHO vaccine. Deo says that because the source of the CRI vaccine is a cultivable microorganism, it is simpler and cheaper to produce than the WHO vaccine, which relies on relatively expensive armadillos and an elaborate purification system. Critics of the WHO proposals are also adamant that a trial should not be started in India until the results of the Malawi and Venezuela studies are known.

The director of the Jalma institute, Dr H. Srinivasan, says that India will lose precious time by relying exclusively on the CRI vaccine. He contends that the WHO vaccine should be tried irrespective of the results from other countries, because of the variation in genetic make-up of the different populations.

The issue is further complicated by a third potential vaccine, developed by Dr G. P. Talwar, at the National Institute of Immunology. Talwar's vaccine, also derived from a cultivable mycobacterium, has been approved for human trial, though not yet field-tested. It is currently being used on 200 patients in two Delhi hospitals to evaluate its therapeutic — but not prophylactic — value.

Dr M. D. Gupte, Jalma's deputy director and a member of the WHO steering committee, says that the WHO is prepared to study all three vaccines in a single, randomized trial. "If all of them are equally effective we will naturally opt for the Indian vaccine."

Talwar is agreeable, but Deo, who still has reservations about the cost and safety of the WHO vaccine (see *Nature* 328, 660; 1987 and 330, 690; 1987), has so far not consented.

Meanwhile, the Central Drug Research Institute in Lucknow has applied to the Drug Controller for a licence to test a fourth vaccine, also mycobacterium-derived.

The ICMR concedes that the situation is "confused" but has apparently made no moves itself to organize a trial using common protocol and design. In the meantime, Talwar and Deo are conducting their own tests, in the full knowledge that without an independent monitoring body the credibility of their results could be questioned.

K. S. Jayaraman

## OSHA standards

Washington

The US Occupational Safety and Health Administration (OSHA) has set new standards for the exposure of workers to over 400 industrial chemicals. The acceptable level of exposure has been lowered for 234 substances, and standards have been established for the first time for another 168 chemicals. Contrary to the agency's usual practice of evaluating and regulating chemical substances one at a time, the exposure limits were adopted from an existing database maintained by an association of industrial hygienists. OSHA officials maintain that the new standards will reduce the number of deaths arising from exposure to dangerous chemicals in the workplace by up to 500 per year, but labour unions have opposed the across-the-board approach as too lenient.

The new standards will affect 3.7 million workers, and will cost the chemical industry an estimated \$4,000 million to implement.

Carol Ezzell

## New head for LBL



Berkeley

CHARLES Cantor is to leave Columbia University to head the Human Genome Center at Lawrence Berkeley Laboratory. Cantor developed the pulsed-field gel electrophoresis technique for separating very large fragments (up to 1,000 kilobases) of DNA, which promises to be vital to the genome mapping project. His expertise will be an important asset to the centre, which has elected to focus its genome-mapping efforts on the cloning and ordering of very large DNA fragments (see *Nature* 330, 9; 1987).

Marcia Barinaga

## Century of looking up

Berkeley

THE University of California's Lick Observatory celebrated its one hundredth birthday on 1 June. The first observatory to be built on a mountain top, Lick sits at nearly 4,000 feet on Mt Hamilton, near San Jose. Its original 36-inch refractor telescope has made many historic discoveries. Lick has acquired six additional telescopes over the years, of which the largest and most powerful is the Shane 120-inch Reflector. Current projects include the Northern Proper Motion programme, which for more than 40 years has been recording shifts in Milky Way stars, in an effort to understand the Galaxy's rotation.

Marcia Barinaga



# The US military's environmental record under fire from Congress

## Washington

THE United States handling of hazardous and radioactive waste at its federal facilities has come under sharp attack from two different quarters.

A congressional subcommittee last week issued a report citing a long list of violations of federal hazardous waste laws by the US government at its Department of Defense (DoD) and Department of Energy (DoE) installations. A separate two-year study by a private organization describes a "massive radioactive contamination crisis" at DoE facilities which produce nuclear weapons.

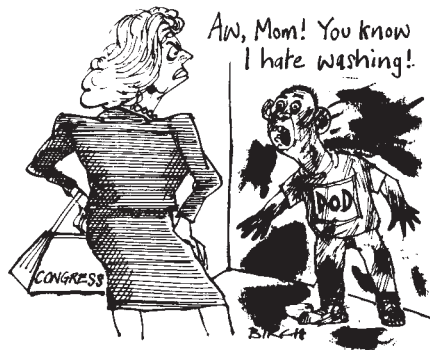
The congressional report highlights several examples of federal noncompliance with environmental regulations. Among these are the Army's Picatinny Arsenal in New Jersey where trichloroethylene (TCE) has been found in groundwater at concentrations 5,000 times greater than the maximum acceptable concentrations set by the Environmental Protection Agency (EPA).

Also cited by the report is the McClellan Air Force Base in Sacramento, California, where similarly high levels of TCE as well as unacceptable levels of arsenic, barium, cadmium, chromium and lead have been found in a municipal well

system that serves 23,000 people in the area. Twelve nearby wells have so far been shut.

In addition to finding high levels of contamination, the report strongly criticizes the Defense and Energy Departments for their failure to comply with EPA clean-up orders. An example is the case of an Army installation in New Mexico, where nearly three years have elapsed with little or no action since clean-up was mandated.

Representative John Dingell (Democrat, Michigan), who chairs the House subcommittee that released the report, calls military facilities "among the worst violators of our hazardous waste laws."



U.S. MILITARY — SHULMAN

and adds that "the [Defense] Department's attitude varies between reluctant compliance and active disregard for the law".

As the congressional report notes, EPA does not have the same influence on federal agencies as on private companies. Although the EPA's enforcement problems often involve Defense Department facilities, the report found similar non-compliance at DoE installations. For example, the Energy Department's nuclear fuel processing centre in Fernald, Ohio, is said to dump 109 million gallons of highly radioactive wastes into storm sewers illegally every year.

The report found that four years after uncovering such serious violations, DoE has yet to finish installing a groundwater monitoring system to analyse the extent of the problem.

Blatant violations of environmental laws at Energy Department nuclear facilities were the specific focus of a second report\* issued last week by the Radioactive Waste Campaign, an environmental group based in New York. The group's study claims to be the first independent analysis of the environmental impact of the entire nuclear weapons complex.

Among the reported findings are that

there are serious levels of radioactive pollution at all 16 of the major nuclear weapons production facilities in the United States, and that the total clean-up cost may be even greater than the Energy Department's estimate of \$100,000 million.

Seth Shulman

\* *Deadly Defense: Military Radioactive Landfills*, Radioactive Waste Campaign, New York, June 1988.

## Ariane-4 launch finally to go ahead after setbacks

### Paris

AFTER a week of setbacks and last-minute repairs, the demonstration flight of Europe's new generation Ariane-4 commercial rocket launcher was scheduled to take place on 15 June at around noon GMT. Originally scheduled to lift off on 8 June, the prototype Ariane-4 rocket developed problems with an on-board computer in the vehicle equipment bay and with the interface between payload and launcher. On-site wiring repairs solved the interface problem, but parts of the computer had to be dismantled and sent to the manufacturers in France and Sweden. A new computer has now been installed and tests carried out at the launch pad in Kourou, French Guyana, are said by Arianespace — the European consortium operating Ariane — to have been "completely satisfactory".

Ariane-4 is to be the 'workhorse' of Europe's commercial launches until the end of the century. With a backlog of 41 satellites, worth \$2,290 million to Arianespace, the company needs Ariane-4 to come into service if it is to keep to its schedule of 8 launches per year. The advantage of Ariane-4 over its predecessors is that it can carry heavier and more bulky payloads, permitting several satellites to be launched at once.

Arianespace won 50 per cent of world commercial satellite launch contracts when the US Challenger shuttle failed, only to find its own rockets grounded for over a year after a third-stage motor failure. With US commercial launches shortly to resume and with the Soviet Union, China and Japan eager to claim a share, Europe needs the demonstration flight of Ariane-4 to go ahead without problems. But Arianespace is confident — in addition to twenty Ariane-4 rockets currently being built, a further fifty have been ordered in order to cope with the consortium's target of 100 satellite launches from now to the year 2000.

Flight 401 will carry a payload of three satellites — a European weather satellite, Meteosat P2, and a US telecommunications satellite, Pan American Satellite-1, both to be put into geostationary orbit, and an amateur radio satellite, Amsat IIC, to be put into an elliptical orbit. Peter Coles

## Europe wants more nuclear glasnost

### London

THE European Commission is increasing its pressure on member states to make information on their nuclear activities more widely available to the public to enable better planning in the event of an accident. The commission has proposed extending to the nuclear sector the principle already established for accident hazards at chemical plants, under which specific information is made available to the public on a regular basis, including details of any emergency plans that exist, and a register of hazards is maintained by the commission. Any similar measures applying to the nuclear sector would be introduced under the legislation on basic standards contained in the Euratom treaty.

The commission also wants to see a number of routine steps to be taken in the event of an emergency, including automatic notification of other member states likely to be affected.

The commissioner responsible for nuclear safety, Mr Stanley Clinton-Davies, hopes the new measures, if adopted, will go some way towards remedying the "woefully inadequate" information now available.

Simon Hadlington



# More about the lion and its claw

SIR—Faulkner's letter<sup>1</sup> on the translation of the famous phrase "*tanquam ex ungue leonem*" (I recognize the lion by its claw), said to have been uttered by John Bernoulli about a work of Newton's, can be bolstered by appeal to older sources. I have a Latin dictionary published in Cambridge in 1693, *Linguae Romanae Dictionarium Luculentum Novum, A New Dictionary in Five Alphabets*. It is not impossible that this is the very dictionary that Newton himself consulted when he sought to probe the nuances of Bernoulli's remark. Here I find, in the relevant one of the five sections, "*Unguis*, . . . A nail of the fingers or toes in man, bird or beast; a claw, a talon: also the hoof of an ox or cow . . ." Thus there are grounds here for preferring, as Faulkner does, "claw" for "paw" and the following word in this dictionary could suggest that "*ungue*" could have an active connotation as well: "*Ungula* . . . The hoof of an horse or other beast: also an hook that is used for the drawing out of a child that is dead in the mother's womb; a tormenting iron, wherewith the sides of malefactors were pinched or burnt." Thus Newton's view of the phrase may well have been like Faulkner's, that Bernoulli knew he had been burned!

The letter does suffer from two obvious errors, however. The title ("Darwin's 'claw', not 'paw'!") is a howler, but more revealing about *Nature's* copyeditors than the letter's author. I suppose there are some dimensions (other than length of name) where it would be proper to equate Darwin with Newton, but surely not in the realm of quotation. The second grievous error is more common, and may be worth attempting a public correction. To misspell Bernoulli as the letter does, with an extra "i" ("Bernouilli"), is not only an affront to the largest important clan in the history of the mathematical sciences, but to the memory of one of the greatest scholars and writers in the pantheon of creative mathematicians. I offer the following quotation from an article about the mathematician Augustus De Morgan, in the original Palgrave.<sup>2</sup>

He has lost the sight of one eye in infancy. The strain upon the remaining eye, involved by his constant industry in accumulating stores of knowledge, must have been severe. The writer once had occasion, in the letter to him, to mention one of the Bernoulli family of mathematicians of the last century, and spelt the name wrongly, that is with two i's, Bernouilli. "Oh," replied De Morgan, "you have deeply offended me. Pray always keep in mind the personal interest I take in

one-eyed philosophers."

STEPHEN M. STIGLER

Department of Statistics,  
University of Chicago,  
5734 University Avenue,  
Chicago,  
Illinois 60637, USA

1. Faulkner, J. *Nature* 332, 200 (1988).

2. Palgrave, R. H. I. (ed.) *Dictionary of Political Economy*, 2. 820 (Macmillan, London, 1896).

SIR—As master of the college in which Newton lived and worked for the greater part of his life, I was intrigued by the speculations of John Faulkner (*Nature* 332, 200; 1988) on possible meanings to be read into the phrase "*ex ungue leonem*" used by Bernoulli about Newton's solution of the brachistochrone problem that he had set as a challenge. I did however remain sceptical because it seemed to me at least possible that the phrase was merely a tag from some classical author, familiar to educated people three centuries ago but no longer current; a pointer in this direction was that "*ex ungue leonem*" scans as the end of an epic or elegiac hexameter. I therefore looked for it in concordances of Horace, Virgil, Ovid, Juvenal and Catullus, but without success. I next asked for help from my friend and colleague, E.W. Handley, Regius Professor of Greek and Fellow of this college, and he promptly found the solution, which he set out in the following letter to me:

Dear Andrew,

Thank you for letting me in on the intriguing little problem of the lion's claw.

Bernoulli is alluding to a proverbial phrase, as you suggest. Faulkner's interpretation seems to me improbable in itself; and I would share your view that the presence of the allusion is against it; for while one can twist a proverb, one would want a much clearer sign from the context than we have in order to suppose that that is being done here.

The proverb is Greek in origin, and is one of a number which relate to popular experiences of inference. 'Painting' or 'recognizing' the lion from its claw (or claws) means basically that if you have that bit (or those bits) you can tell that it came from an animal of a certain kind and size: a single sample of Newton's work makes it possible to be clear what kind and size of a mathematician he is.

In detail (as often happens) the story is more intricate. Like you, I began by looking for a Latin tag from the end of a hexameter, or a proverb in the traditional form of a paroemiac, which metrically comes to the same thing, but does not require one to postulate a poem. But there is no trace of such a proverb in the standard work on the subject, Otto's *Sprichwörter der Römer* (1890), nor indeed in the list of 'lion' proverbs in the *Thesaurus Linguae Latinae* s.v. leo; though I am bound to say that I was somewhat relieved that you had checked the likely poets in modern concordances. This is a dead end, so far as I can see.

In Greek the story is different. Various dif-

ferent treatises in the *Corpus Paroemiographorum Graecorum* list ἐκ τοῦ ὀνύχων τὸν λεόντα *vel sim.*, for instance Diogenianus v.15 (vol.i, p.252 Leutsch-Schneidewin) with an explanation like the one given above; and from the editors' notes there one gathers that the earliest attestation is in Alcaeus (vii-vi B.C.) as quoted by Plutarch, *de defectu oraculorum* 3 (*Moralia* 410 C). It might well be that that is Bernoulli's source. I had wondered about Erasmus, *Adagia* or a derivative, but at *Adagia* 834 (I.ix.34) E. offers *ex unguibus aestimare leonem* as his version of the Greek, with plural claws; while Plutarch gives us both the singular 'claw' and a work with some mathematical content which might have made it memorable to Bernoulli—but the *Moralia* were in any case quite widely read at the time. It is also true that the standard Latin translation of Plutarch's phrase produces—by literal means—*ex ungue leonem* out of ἐξ ὀνυχός τον λεόντα, thus creating the appearance of a tag or proverb in Latin which led to our wild goose chase.

Yours sincerely,

(sgd) Eric Handley

I am most grateful to Professor Handley for searching out this solution to the problem, and for allowing me to publish his letter.

ANDREW HUXLEY

The Master's Lodge,  
Trinity College,  
Cambridge CB2 1TQ, UK

SIR—John Faulkner<sup>1</sup> interpreted the Latin phrase "*tanquam ex ungue leonem*" as a sign of contempt, "the old lion has given me the finger". The old lion would be Isaac Newton, the person that made the remark Johann Bernoulli.

Fortunately, Bernoulli knew his classics better than Faulkner, for the well-known phrase "*ex ungue leonem*" is a quotation from the Greek poet Alcaeus<sup>2</sup>. This phrase was made more or less famous by the Dutch humanist Erasmus who, in his *Adagia* (*Proverbs*), quoted it as '*leonem ex unguibus aestimare*': 'to judge the lion by the claws', that is, by a slight but characteristic mark.

Erasmus had some doubt about the source of this proverb. In my opinion this is why other books of quotation do not mention it. Among humanists this phrase was well-known. I myself discovered it for instance in the *Apotheosis of Macropedius* (1565), a poem by Macropedius' pupil Vladeraccus<sup>3</sup>. No contempt, no visceral feelings, but admiration. Quoting such a phrase, Bernoulli showed his admiration: "You shall know Newton by his fruits".

JAN BLOEMENDAL

Bachlaan 27,  
3781 HC Voorthuizen,  
The Netherlands

1. *Nature* 332, 200 (1988).

2. fr. 113 (Bergk): ἐξ ὀνυχός τον λεόντα (sc. τεκμαίρεσθαι).

3. Chr. Vladeraccus, *Apotheosis Macropedii*, vss. 145ff. *Unquibus ut magni magna virtute leones Noscuntur, volucres prodit et ipse color, Sic quoque discipuli monstrant virtute magistrum.* . . . 'As great lions with great courage are known by their claws and as their colour betrays the birds, so do his pupils by their virtue show the master.'

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only. □

# Quasicrystals soldier on?

*Attempts continue to account for the occurrence of quasicrystals with fivefold symmetry, and are to be applauded. The case for believing them important is less sure.*

THERE seems still to be life in the problem of quasicrystals, and of whether they are explicable. The essence of the problem, it will be recalled, is that fivefold symmetry is flatly forbidden for a crystal lattice in which the crystallizing units have both an intrinsic symmetry and the translational symmetry of a crystal lattice as well. Yet, as the world knows, we have been living with exceptions to that rule for the best part of three years. Apart from the original intermetallic compound of Al and Mn, there are now ternary compounds such as Al-Cu-Li and Ga-Mg-Zn which have phases showing distinct icosahedral symmetry.

Not that explanations are impossible. The earliest were essentially generalizations of the way in which it is possible completely to fill a plane surface by a suitable arrangement of tiles of two distinct but related rhombi — the arrangement called the Penrose tiling. The ratio of the major and minor diagonals of each rhombus is the golden mean, also the natural number that crops up in the definition of a Fibonacci series.

This tiling is not a two-dimensional crystal, for there is nothing corresponding to a unit cell, nor to directions in two-dimensional space that would correspond to crystal axes, but that is an academic question provided that the arrangement can be essentially infinite. Then, there is no reason why such a tiling of the plane should not represent a two-dimensional solid of some kind, whence the term "quasicrystal". By now, there are three and even many-dimensional generalizations of the Penrose tiling, and thus models for the existence of three-dimensional quasicrystals. There is no doubt that some of the elements of these structures do, indeed, have fivefold symmetry. The interesting question is whether these structures are those of the solids that actually exist.

Observationally, the issue is clouded by the difficulty of making structural measurements with materials that are usually made at high temperatures, and by the likelihood that the conditions in which most specimens are made are far from equilibrium.

On balance, three distinct positions are recognizable in the range of opinion on offer — quasicrystals which are infinite in extent do indeed exist; quasicrystals of infinite extent are possible, but in reality are so shot-through with dislocations and

other internal defects that quasicrystallinity is limited to a few hundred interatomic distances; and, on the radical (or is it "conservative"?) extreme, Pauling's view that the structures called quasicrystals are really solid materials formed by the interspersing of microscopic crystallographic twins.

One way of telling which of these possibilities, or which combination of them, may arise in practice is to concentrate on gross features of the actual specimens made in the laboratory, asking simple questions such as "If they have fivefold symmetry on an atomic scale, should they not also have faces, or facets, with fivefold symmetry?" This problem is first that of calculating the equilibrium shape of the facets of a quasicrystal (which will be a function of the temperature). A diagnostic test that bears on this is whether the facets of a small quasicrystal should be rough or smooth, on a scale measured by a few interatomic spacings, for this will determine how readily atoms will migrate to their equilibrium positions within a structure.

Richard Lipowsky (from the nuclear research centre at Jülich in West Germany) and Christopher L. Henley (Boston University) have made an interesting attack on the second problem (*Phys. Rev. Lett.* **60**, 2394; 1988). The issue is that of what the most likely (Boltzmann-weighted) configuration of a surface will be in a quasicrystal. The calculation is simply that of the enumeration of the numbers of bonds between neighbouring Penrose-like tiles broken as surfaces of different configuration are formed. The equilibrium is determined by the increase of energy on breaking the bonds and the loss of entropy that results.

What emerges is that the stability of a surface formed by breaking apart a quasicrystal is sharply different in the two- and the three-dimensional cases. The energy term is an interfacial term, but the entropy is a function of the dimensions of the bulk. Lipowsky and Henley accordingly conclude that two-dimensional quasicrystals will have decagonal shape at absolute zero, but will become rounded at the corners as the temperature is increased.

Three-dimensional crystals are a different kettle of fish, because the relatively greater importance of the entropy will keep decagonal faces decagonal for longer. By the same test, of course, badly

formed quasicrystals with no recognizable fivefold symmetry on the outside will convert to forms in which the internal symmetry is externalized as they are annealed. Indeed, the message here is that the way to tell the internal symmetry of a quasicrystal is to anneal it, when its internal symmetry will be revealed.

The same issue of the same journal has a more recondite quarrel on a related topic. Kevin Ingersent and Paul J. Steinhardt (co-discoverer of the Al-Mn quasicrystalline material) have obviously been belatedly cut to the quick by an earlier paper (Ho *et al.* *Phys. Rev. Lett.* **59**, 1116; 1987) asserting that a model of a quasicrystal in which the forces between atoms are strictly attractive leads to the conclusions that decagonal faces cannot exist.

"So what?", Ingersent and Steinhardt stoutly protest (*Phys. Rev. Lett.* **60**, 2444; 1988). They construct a model allowing for the surface energy of a quasicrystal, and wrapping the entropy term within a variable called the "surface free energy". The interest of this brief note is a sequence of diagrams showing the shapes of the quasicrystals that will arise with different proportions of repulsion and attraction. The repulsive component of the interatomic force, they say, is crucial: the greater its relative amount, the more faceted and less spherical the quasicrystallites become.

Lipowsky and Henley have made essentially the same point; perhaps the best way to tell the structure of a quasicrystallite is to make a few specimens (usually a few micrometres across) and to anneal them. Then the internal structures will be externalized. Maybe the experimentalists in the field will be sufficiently challenged to try.

Meanwhile, we have all learned a great deal about what may be called, not unreasonably unfairly, the great quasicrystal ramp (which will continue). First, there is such a thing as dodecahedral (fivefold) symmetry, at least on the scale of atoms in a three-dimensional atomic array. Second, there are ways of making (or at least modelling) three-dimensional arrays of atoms which hold together and may be infinite, or at least large, in extent. Third, so little has been discovered, in the past three years, from the properties of quasicrystals about the nature of the solid state that the whole episode may have served merely to sharpen people's teeth for more interesting battles.

John Maddox



## Oceanography

## Submarine salt lenses

John Marshall

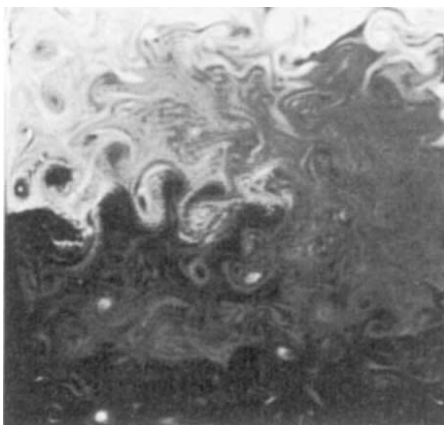
LAURENCE Armi and collaborators report, on page 649 of this issue<sup>1</sup>, a unique series of oceanographic observations in which, over a period of 2 years, they document the life history of a subsurface coherent lens of salty water formed from outflow through the Strait of Gibraltar. Meddies, as they are called, have a vertical extent of about 900 metres and are typically 50 kilometres in diameter<sup>2</sup>. Spinning in the opposite sense to the Earth, particles of fluid circle the axis of the Meddy in about a week. They have a lifetime of perhaps 2 years and have been detected drifting at a depth of 1 kilometre as far west as the Bahamas. The experiment suggests a rather fascinating mechanism by which the salty water of the Mediterranean finds its way into the North Atlantic. Rather than the salty influence being diffused or advected by large-scale circulation into the interior, it can be transported in concentrated spinning blobs. The salt 'tongue', spreading westwards from Gibraltar to beyond the mid-Atlantic ridge, could be formed from the decay of Meddies releasing their salty contents.

Imagine the Mediterranean outflow as a giant submerged tap dripping into the North Atlantic. The blobs of fluid settle at their neutrally buoyant level of about 1 kilometre. The volume and salinity of observed Meddies suggest that a birth rate of one every 10 days would be sufficient to carry the entire outflow of salty water. With a life expectancy of 1–2 years, there could be 50 Meddies existing at any one time in the North Atlantic.

The Meddies carry with them properties of their formation region — the weakly stratified, salty waters of the Mediterranean outflow. The single most important factor controlling their dynamics is not that they are salty, but that they are tagged by anomalously low values of the dynamically active tracer, potential vorticity. In the present context the potential vorticity  $PV$  can be defined as  $(\zeta + f)/h$ :  $f$  is the planetary vorticity, the Coriolis parameter measuring the Earth's spin in the direction of the local vertical;  $\zeta$  is the relative vorticity measuring the spin of the vortex relative to the Earth; and  $h$  is the vertical distance between surfaces of constant density (isopycnals), a measure of the stratification. Convective overturning of the upper layers of the Mediterranean destroys the stratification, forming a reservoir of low- $PV$  water which drips into the North Atlantic. In adiabatic, inviscid motion,  $PV$  is conserved following the motion like any other fixed feature of a parcel of fluid. Thus, the blobs of

water remember the small values of  $PV$  typical of the Mediterranean reservoir.

A further crucial dynamical factor is that the Meddy is in balanced motion; Coriolis forces acting on the spinning vortex are balanced by radial pressure gradients (it is in geostrophic balance) and gravity is balanced by vertical pressure gradients (hydrostatic balance). In such a balanced vortex, anomalously low  $PV$



Eddies and rings at a depth of 400 metres in a numerical simulation of idealized ocean gyres as seen through the potential vorticity field; dark (light) indicates fluid of low (high)  $PV$  spinning in the opposite (same) sense as the Earth. The model is driven by an imposed wind stress at its upper surface, and has two counter-rotating gyres in a 3,000-kilometre-square basin. At their confluence in mid-basin, eddies and rings are formed through the instability of the internal jet (the model's Gulf Stream). The typical eddy scale is set by the 'Rossby' radius of deformation, chosen here to be 50 kilometres. Dark coherent blobs of very-low- $PV$  fluid can readily be seen and dynamically have much in common with Meddies. White blobs are their mirror images.

must manifest itself both as a weakening of the stratification and as anticyclonic relative vorticity<sup>3</sup> (opposite in sign to the spin of the Earth). Observed horizontal scales, on or near the 'Rossby' radius of 50 kilometres, suggest that the anomaly in  $PV$  is equally partitioned between relative vorticity and stretching. Thus the Meddy appears as a double-convex lens (a thickening of isopycnal layers) in anti-cyclonic motion.

The coherence and persistence of Meddies arises from the fact that their  $PV$  is so anomalous that a closed circulation is induced which traps the fluid inside it; closed  $PV$  contours are the walls which allow the Meddy to carry its anomalous contents large distances (including the neutrally buoyant floats used to track them) and form the deep-water laboratory studied by Armi and collaborators<sup>1</sup>.

The vertical influence of the Meddy's  $PV$  anomaly will decay over the Rossby height scale  $fL/N$ , where  $L$  is the horizontal length scale of the anomaly and  $N$  is a measure of the static stability. In the upper mid-thermocline,  $N/f \approx 50$ , suggesting that they have a vertical extent of perhaps 1 kilometre, if  $L$  is taken to be 50 kilometres. It is likely, then, that the Meddy induces flow at the surface (not the bottom) but that the influence is weak and unlikely to be strong enough to permit trapping of fluid properties at the surface. Meddies are thus difficult (if not impossible) to detect remotely in, say, sea-surface temperature or ocean colour imagery.

So the blob of low- $PV$  water floats serenely at mid-depth, isolated from frictional torques associated with surface and bottom boundary layers. But this state of affairs cannot persist indefinitely. At the edge of the blob there are strong  $PV$  gradients on which small-scale instabilities can feed; Armi *et al.*<sup>1</sup> find evidence of both lateral and vertical mixing (thermocline intrusions and salt fingering) and high levels of inertial activity. Radiation of energy through the agency of Rossby waves (horizontal undulations of the ambient  $PV$  field) is also a possible decay mechanism. In the face of these dissipative processes, it is remarkable that the vortex can survive for so long.

An interesting question is raised by the observed track of the Meddy; why did it move southwards, rather than, say, westwards as might have been expected<sup>4</sup>? There is no clear explanation, but perhaps the large-scale geometry of the  $PV$  field is the clue. The  $PV$  contours are the free paths, the rails along which information (and Meddies) travel with least resistance. In textbook Rossby-wave theory, the large-scale  $PV$  gradient is set by the meridional variation in  $f$ , the so-called  $\beta$ -effect, and so information from the east travels westwards. But perhaps here we should reinterpret the theory as meaning 'pseudo-westward' along the free paths set by the large-scale  $PV$  field. At mid-depth in the thermocline, they depart from latitude circles and have a significant north-south component associated with the subtropical gyre of the North Atlantic<sup>5</sup>. So one possible interpretation of the observed path of the Meddy is that it is propagating along the  $PV$  rails induced by the gyre.

What are the wider implications of the study? Evidence from high-resolution numerical models of eddy-rich ocean gyres (see figure) suggest that coherent vortices are a ubiquitous feature of the oceans. Meddies are just a spectacular example of one of the canonical forms; their mirror images, double-concave lenses in cyclonic motion are associated with fluid of anomalously high  $PV$ . Cyclonic and anticyclonic coherent

structures will be formed wherever there are instabilities feeding off large PV gradients at the interface between water masses of very different PV characteristics. Warm and cold rings, for example, are formed through a hydrodynamical instability of the Gulf Stream marking the frontal region between the high-PV slope waters to its north and the low-PV waters of the Sargasso Sea to its south (see figure). Warm rings are coherent vortices of low-PV water which are dynamically similar to Meddies. But unlike Meddies they are somewhat shallower, readily eroded by active surface processes and hence shorter lived.

The problem for the dynamical oceanographer is that our ways of thinking and mathematical paradigms are too often

couched in terms of linearized studies of wave propagation and normal mode instability analyses, and so are ill-equipped to handle a fluid populated by highly non-linear coherent vortices. Studies such as the one described by Armi *et al.* remind us that perhaps we should take more notice of those theories that emphasize the particle rather than the wave-like aspects of geophysical fluids. □

1. Armi, L. *et al.* *Nature* **333**, 649–651 (1988).
2. Armi, L. & Zenk, W. *J. phys. Oceanog.* **14**, 1560–1676 (1984).
3. Hoskins, B.J., McIntyre, M.E. & Robertson, A.W. *Q. J. R. met. Soc.*, **3**, 877–946 (1985).
4. Killworth, P.D. *J. phys. Oceanog.* **16**, 709–716 (1986).
5. McDowell, S., Rhines, P. & Keffer, T. *J. phys. Oceanog.* **12**, 1417–1436 (1982).

John Marshall is in the Department of Physics, Imperial College of Science and Technology, London SW7 2BZ, UK.

## Nitrogen fixation

# A third bacterial nitrogenase

R. Cammack

NITROGENASE is the enzyme in bacteria responsible for fixing atmospheric nitrogen to ammonia. When isolated from various microorganisms, it was found to be an iron-sulphur protein containing molybdenum. Microorganisms could not fix nitrogen if molybdenum was excluded from the growth medium. Actually, this was not quite true; with some organisms it was extremely difficult to prevent nitrogen fixation entirely. This was attributed to the extraordinary ability of the bacteria to scavenge traces of molybdenum from the culture media and vessels. Two years

fixation by bacteria.

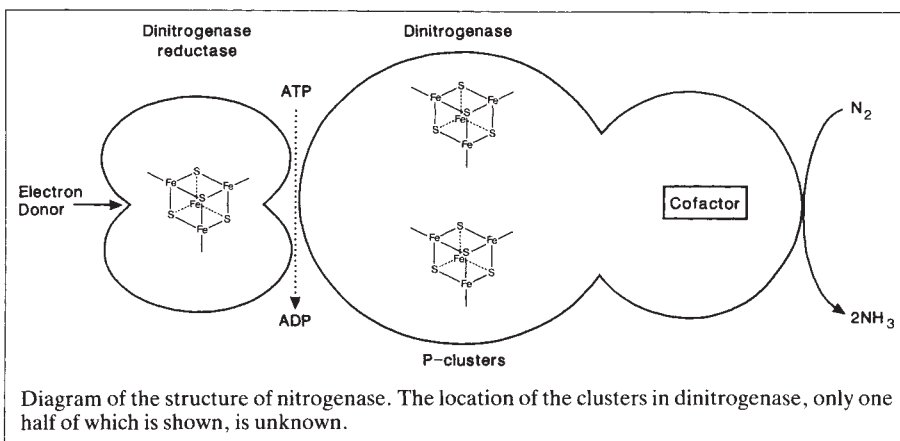
The existence of alternative nitrogenases had been suspected for many years. Convincing evidence for this idea was difficult to obtain because it turned out that biosynthesis of the alternative nitrogenase is strongly repressed by molybdenum in the growth medium. Moreover, the alternative nitrogenase is not very active in the reduction of acetylene to ethylene, a reaction which had become the standard method for detecting nitrogenase activity<sup>6</sup>.

The existence of one alternative nitro-

amounts<sup>2,4</sup>, and is probably responsible for the earlier observations<sup>9</sup> of nitrogen fixation by *A. vinelandii* grown on molybdenum-free medium. Eady and collaborators at the AFRC Nitrogen Fixation Laboratory at Sussex University have carried the genetic experiment one stage further<sup>1</sup>, also deleting the genes *nifHDK*\* encoding the V-nitrogenase. Sure enough, the bacteria could still fix nitrogen, and a third nitrogenase could be extracted.

Such substitutions of one metal for another are unusual in enzyme systems: cells are generally specific in inserting the 'right' element. Although metal ions involved in catalysis can sometimes be removed artificially from enzymes and replaced with other elements, the effect on catalytic activity is considerable. In this case, the molybdenum-, vanadium- and iron-containing nitrogenases are all different proteins, although they have a similar structure.

The new iron-containing enzyme (Fe-nitrogenase), the V-nitrogenase and the well-established Mo-nitrogenase, are all similar in that they each comprise a large protein, sometimes called dinitrogenase, where the nitrogenase is actually fixed, together with a smaller iron-sulphur protein, termed dinitrogenase reductase (see figure). The Fe-nitrogenase and the V-nitrogenase differ slightly in that they seem to contain a small third subunit whose function is unknown<sup>2,5</sup>. The amino-acid sequences of the three types of nitrogenase are very similar<sup>3,4</sup>. To emphasize this similarity, it has been possible to extract the inorganic cofactors, termed Fe-Mo-co and Fe-Va-co, from the Mo- and V-dinitrogenases, respectively. These cofactors are similar in their spectroscopic properties, and can restore catalytic activity to a mutant Mo-nitrogenase which lacks the cofactor<sup>10</sup>. When Fe-Va-co is inserted, the mutant enzyme acquires the ability to reduce ethylene to ethane, which is the property of the V- and not the Mo-nitrogenase. This observation supports the view that it is the cofactor which is the site of nitrogen fixation in the enzyme. The properties of the putative



ago, it was found that strains of *Azotobacter vinelandii* and *A. chroococcum* could produce an alternative nitrogenase containing vanadium instead of molybdenum (see my News and Views article<sup>1</sup>). Now, a third nitrogenase has been isolated from *A. vinelandii* which contains neither element and appears to use iron instead<sup>2,3</sup>. These results, which were presented at a recent symposium in Cologne<sup>3,5</sup>, overturn the accepted view that molybdenum is an essential trace element for nitrogen

genase was demonstrated genetically by deleting the structural genes, *nifHDK*, encoding the conventional molybdenum-nitrogenase, from *Azotobacter* species, and isolating an enzyme (V-nitrogenase), which contained stoichiometric amounts of vanadium and only trace amounts of molybdenum<sup>4,6-8</sup>. P. E. Bishop's group in North Carolina State University has now isolated another nitrogenase from such a deletion strain. This protein contains iron as the only metal ion in significant

1. Cammack, R. *Nature* **322**, 312 (1986).
2. Chisnell, J.R., Premakumar, R. & Bishop, P.E. *J. Bact.* **170**, 27–33 (1988).
3. Pau, R.N., Mitchenall, L.A., Eady, R.R. & Robson, R.L. in *Proc. 7th Int. Symp. Nitrogen Fixation, Cologne* (eds Bothe, H. & Newton, W.E.) (Fisher, Stuttgart, in the press).
4. Bishop, P.E. *et al.* in *Proc. 7th Int. Symp. Nitrogen Fixation, Cologne* (eds Bothe, H. & Newton, W.E.) (Fisher, Stuttgart, in the press).
5. Eady, R.R. *et al.* in *Proc. 7th Int. Symp. Nitrogen Fixation, Cologne* (eds Bothe, H. & Newton, W.E.) (Fisher, Stuttgart, in the press).
6. Dilworth, M.J., Eady, R.R., Robson, R.L. & Miller, R.W. *Nature* **327**, 167–169 (1987).
7. Robson, R.L. *et al.* *Nature* **322**, 388–396 (1986).
8. Hales, B.J., Case, E.E., Morningstar, J. E., Dzeda, M.F. & Maeterer, L.A. *Biochemistry* **25**, 7251–7255 (1986).
9. Bishop, P.E., Jarlenski, D.M.L. & Hetherington, D.R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7342–7346 (1980).
10. Smith, B.E., Eady, R.R., Lowe, D.J. & Gormal, D. *Biochem. J.* **250**, 299–302 (1988).
11. Hardy, R.W.F., Bottomley, F. & Burns, R.C. *A Treatise on Dinitrogen Fixation* Vol. 1, 172 (Wiley, Chichester, 1979).



cofactor, Fe-Fe-co, of the Fe-nitrogenase will be of great interest.

So far the Fe-nitrogenase has been found only in *A. vinelandii* and not in the other organisms used in studies of the Mo-nitrogenase, such as *A. chroococcum* or *Klebsiella pneumoniae*. The Fe-nitrogenase is repressed if either molybdenum or vanadium is present in the growth medium<sup>4</sup>. A possible reason is that the isolated enzyme is less efficient in nitrogen fixation, producing more hydrogen in a wasteful side-reaction.

These recent developments show that molybdenum or vanadium ion is not as

essential to the activity of the enzyme as had previously been supposed and that it is the iron in the cofactor that is responsible for binding dinitrogen. Note that the Haber process, used industrially for fixing nitrogen to ammonia, uses an iron catalyst, although this is metallic and uses various promoters<sup>11</sup>. The search for inorganic complexes that mimic the activity of nitrogenase may have focused too closely on molybdenum complexes. □

R. Cammack is a Reader in Biochemistry in the Division of Biomolecular Sciences, King's College, London W8 7AH, UK.

## Stellar nucleosynthesis

# Neutron reaction rates revised

Roger J. Tayler

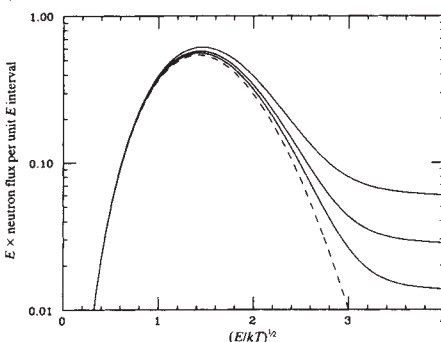
CALCULATIONS of nuclear reactions in stars are usually made with the assumption that the distributions of velocities of the interacting particles are those appropriate to thermodynamic equilibrium at the local kinetic temperature. In two recent papers, Malaney<sup>1</sup> and Petrov and Shlyakhter<sup>2</sup> argue that in the late stages of stellar evolution, when the stellar core consists primarily of nuclei such as <sup>28</sup>Si, the distribution of neutron velocities can deviate from thermodynamic equilibrium and that as a result there can be order-of-magnitude revisions in the rates of some neutron-induced reactions.

The assumption that conditions in the deep interiors of stars are very close to thermodynamic equilibrium means that, provided the matter is non-degenerate, all particles have a maxwellian distribution of velocities and the photons have a Planck distribution of energies. In a region in which no nuclear reactions are currently occurring, the assumption of thermodynamic equilibrium is likely to be sound because the timescales of all processes leading to equilibrium are much shorter than the stellar-evolution timescale.

The assumption needs to be reconsidered for matter in the presence of nuclear reactions, because both particles and photons emitted in nuclear reactions tend to possess energies which are greatly in excess of the typical thermal energies. For example, even in hydrogen burning in stars the emitted particles have energies that are measured in MeV, whereas thermal energies are little more than about 1 keV.

It is usually expected that the strongly and electromagnetically interacting supra-thermal particles will rapidly share their energies with the rest of the matter before they can initiate any reactions with rates different from those appropriate to the local temperature. In contrast, the weakly interacting neutrinos created usually

escape freely from the star carrying their excess energy with them. Clayton<sup>3</sup> suggested, in an attempt to resolve the solar-neutrino problem, that the nuclei in the Sun might not in fact have a maxwellian distribution. Nuclear reactions slightly out of thermodynamic equilibrium could produce a modified spectrum of neutrino



Neutron energy distributions for different values of Malaney's parameter  $\Delta \approx 2A\sigma_n/\sigma_s$ , where  $A$  is the target nuclear mass number, and  $\sigma_n$  and  $\sigma_s$  are the neutron absorption and scattering cross-sections. Broken line, maxwellian distribution; upper, middle and lower solid lines,  $\Delta = 0.2, 0.1$  and  $0.05$ , respectively. (From ref. 1.)

energies, which would explain why fewer neutrinos than expected were detected by the experiment of Davis and Evans<sup>4</sup>. But this suggestion has not been confirmed.

Malaney<sup>1</sup> and Petrov and Shlyakhter<sup>2</sup> now propose that departures from a maxwellian distribution of velocities are important at a much later stage of stellar evolution, which occurs only in massive stars. The process known as silicon burning is the name for a complex series of nuclear reactions involving silicon and its neighbours in the periodic table which converts them into the nuclei close to iron. Neutrons are released in some of these reactions which in turn induce reactions in other nuclei.

The reactions are of two main types:  $(n, \gamma)$  whereby the neutron is added to the nucleus and a photon is emitted; and  $(n, p)$  in which a proton is emitted. The authors argue that the neutrons can induce reactions before their velocity distribution has relaxed to a maxwellian form, and so modify the products of nucleosynthesis. Malaney uses the arguments developed in a study of neutron thermalization in the cores of nuclear reactors<sup>5</sup> to show that the distortion of the maxwellian energy distribution can be expressed in terms of a quantity  $\Delta$  which depends on the ratio of the absorption and scattering cross-sections (probabilities) for neutrons (see figure).

The rate of  $(n, \gamma)$  reactions is not significantly affected by the shift of neutrons to higher energies but this is not true for  $(n, p)$  reactions. Protons have to overcome a Coulomb electrostatic barrier to get out of the nucleus and this means that the additional neutron energy can produce a large increase in the  $(n, p)$  reaction rate. Petrov and Shlyakhter present a similar discussion<sup>2</sup> and stress that the enhancement of the reaction rate can depend significantly on which reaction releases the neutrons. As a result the reactions in silicon burning will tend, at least initially, to produce rather more neutron-rich nuclei, although the protons will, themselves, interact very quickly.

Neither Malaney nor Petrov and Shlyakhter incorporate the new reaction rates into a full discussion of silicon burning such as that given by Thielemann and Arnett<sup>6</sup>. Until this is done, it will not be easy to see whether there are any important implications for the observed distribution of elements and isotopes. Much of the material that undergoes silicon burning in the quasi-static stages of stellar evolution is subsequently converted to neutrons as a result of electron-capture reactions and the core of the massive star becomes a neutron star. Also, if silicon burning occurs in the course of a supernova explosion, the reaction timescales become so short that out-of-equilibrium effects could be even more important than in quasi-static stages of stellar evolution. What the authors have done is to remind astrophysicists that they must check that the assumption of thermodynamic equilibrium is valid for any process which they are studying. □

1. Malaney, R.A. *Mon. Not. R. astr. Soc.* **231**, 657-662 (1988).
2. Petrov, Yu.V. & Shlyakhter, A.I. *Astrophys. J.* **327**, 294-297 (1988).
3. Clayton, D.D. *Nature* **249**, 131 (1974).
4. Davis, R. Jr & Evans, J.C. in *Proc. XIII Int. Conf. Cosmic Rays 3*, 2001-2006 (University of Denver, 1973).
5. Weinberg, A.M. & Wigner, E.P. *The Physical Theory of Neutron Chain Reactors* (University of Chicago Press, 1958).
6. Thielemann, F.-K. & Arnett, W.D. in *Nucleosynthesis: Challenges and New Developments* (eds Arnett, W.D. & Truran, J.W.) 151-189 (University of Chicago Press, 1985).

Roger J. Tayler is in the Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.

## Population biology

## Mechanisms of coexistence

Wallace Arthur

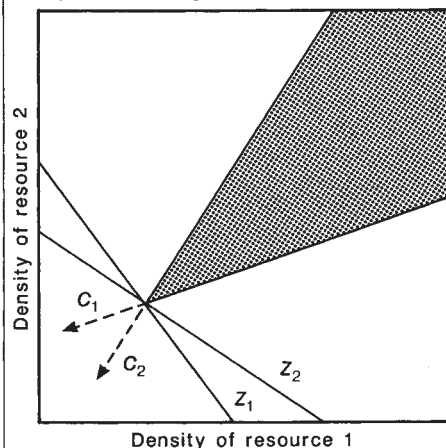
Is interspecific competition an important determinant of major community properties such as species diversity and, if so, what sort of influence does it have? Ecologists have long sought to answer this question, and have approached it in various ways. One aspect of the problem is determining how common competition is in nature in the first place. A diverse array of field experiments<sup>1,2</sup> shows that it is indeed reasonably common, though certainly not universal, among closely related sympatric species. Another aspect of the problem is elucidating the population dynamics of competition, that is, attempting to answer the question of how competition 'works' in those systems in which it is known to be taking place. It is this aspect of the problem that is addressed by Karl Rothhaupt on page 660 of this issue<sup>3</sup>, in a fascinating series of experiments on two competing species of rotifer.

Of course, there have been many previous attempts to elucidate the population dynamics of competition, from the classic work of Gause<sup>4,5</sup> onwards. But many earlier authors failed to investigate the biological mechanisms of competition and coexistence, instead examining only the compatibility of their experimental data with models of the Lotka-Volterra genre which are not, themselves, mechanistically explicit. Others — including Gause himself — postulated mechanisms of coexistence but did not quantify them or carry out the necessary experimental protocol<sup>6</sup> to distinguish the proposed mechanism from possible alternatives. Given this state of affairs, a particularly useful approach in current studies of competition is the comparison of experimentally derived data with models, such as those of Tilman<sup>7</sup>, which are mechanistically explicit in their approach. This is precisely what Rothhaupt sets out to do.

His experimental system involves two closely related species of herbivorous rotifer, *Brachionus rubens* and *B. calyciflorus*, which eat two species of alga — one species of *Chlamydomonas* and one of *Monoraphidium*. Both species of rotifer eat both algal species, but *B. rubens* prefers and does better on *Monoraphidium*, whereas the reverse is true of *B. calyciflorus*. In a four-component mixture containing both consumers and both resources, exploitative competition occurs between the rotifers, and the population dynamics of their competitive interaction can be followed for a sufficient time to determine whether the outcome is stable coexistence or competitive

exclusion. Rothhaupt performs twelve such experiments, each under different resource conditions. In some cases, *B. rubens* competitively excludes its congener; in others the situation is reversed, with *B. calyciflorus* winning; and in others the two rotifers reach a state of stable coexistence.

Rothhaupt then assesses the compatibility of his experimental results with



The division of the resource plane into a zone of predicted coexistence of the consumers (shaded) and a complementary zone in which other outcomes are predicted is achieved by projecting the consumption vectors ( $c_1$  and  $c_2$ ) through the intersection of the zero net growth isoclines ( $z_1$  and  $z_2$ ). In the shaded zone, the joint effect of resource supply and consumption is such as to move the resource densities towards the equilibrium point<sup>7</sup>.

Tilman's mechanistic theory<sup>7</sup> of exploitative competition, where the conditions for stable coexistence are phrased in terms of resource supply and consumption, rather than in terms of abstract 'competition coefficients'. Specifically, Tilman's approach for a two-consumer, two-resource system involves subdivision of a resource plane — a two-dimensional space bounded by the densities of the resource species — into a zone in which coexistence is predicted and a complementary zone in which it is not. The way in which the resource plane is divided into these zones is complex<sup>7</sup>, and I do not have the space to explain it fully here, but the two relevant statistics (see figure) are the consumers' zero-net-growth isoclines (which join up points of zero population growth) and their consumption vectors (which indicate the effect on the resource densities of the consumers' feeding activities). Superimposing his results on to the appropriate resource plane, Rothhaupt finds a high degree of correspondence between prediction and

observation. In five out of six cases where coexistence is predicted, it is found to occur, and in all six cases where competitive exclusion is predicted, one species does indeed exclude the other. Moreover, which species wins in the exclusion cases is also predicted by the theory, and in all cases the 'right' species wins.

Although this is an impressive correspondence between theory and experiment, some questions remain unanswered. Foremost among these is the question of how error-prone measurement of the consumption vectors is under any given set of conditions, and also how sensitive these vectors are to slight changes in extrinsic variables such as temperature. Because the consumption vectors delimit the zone of predicted coexistence, such questions are of crucial importance. The degree of error associated with measurements made under a single set of conditions is important in the interpretation of laboratory studies. It is particularly important in those, such as Rothhaupt's, where some of the resource treatments are perilously close to the edge of the zone of predicted coexistence. The degree of sensitivity of the consumption vectors to changes in extrinsic conditions is relevant to the application of this method to natural populations. If the zone of predicted coexistence alters its size and shape on a timescale less than that on which the competitive interaction normally delivers its end-result (exclusion or stable equilibrium), then considerable difficulties will be encountered. If not, then the method could be applicable to the field situation with no more than the usual additional difficulties associated with the laboratory-to-field transition.

Such difficulties aside, the combination of experimental studies (in both field and laboratory) with mechanistically explicit models provides a promising approach for the continuing elucidation of how interspecific competition 'works'. How useful this elucidation will be in the study of the main features of community structure depends on, among other things, the extent to which the population dynamics of competition in radically different groups of consumers follow the same rules. Many careful experimental studies are necessary before we can discover whether general rules for competition are a reality of nature or a theoretician's dream. □

1. Schoener, T.W. *Am. Nat.* **122**, 240–285 (1983).
2. Connell, J.H. *Am. Nat.* **122**, 661–696 (1983).
3. Rothhaupt, K.O. *Nature* **333**, 660–662 (1988).
4. Gause, G.F. *The Struggle for Existence* (Williams and Wilkins, Baltimore, 1934).
5. Gause, G.F. *Act. Sci. et Ind.* **277** (1935).
6. Arthur, W. *The Niche in Competition and Evolution* (Wiley, Chichester, 1987).
7. Tilman, D. *Resource Competition and Community Structure* (Princeton University Press, 1982).

Wallace Arthur is in the School of Biology, Sunderland Polytechnic, Sunderland SR1 3SD, UK.



## Protein structure

## The 14th barrel rolls out

Cyrus Chothia

In proteins, stretches of the polypeptide chain form  $\alpha$ -helices or strands of  $\beta$ -sheet. These 'secondary structures' then fold up on each other to form the protein's three-dimensional structure. In 1975, the fold of the polypeptide chain in the enzyme triose phosphate isomerase (TIM) was discovered<sup>1</sup> to be particularly simple and elegant. Its secondary structure consists of eight strands of  $\beta$ -sheet alternating with one, or sometimes two,  $\alpha$ -helices. The chain folds so that the eight strands form a parallel  $\beta$ -sheet which twists to give a closed barrel-like structure at the centre of the molecule. The helices that link the parallel strands pack onto the exterior of the barrel (see Fig. 1). The TIM structure was the first known example of the  $\alpha/\beta$ -barrel fold. On page 683 of this issue<sup>2</sup>, Lebioda and Stec, in their description of the structure of yeast enolase, present the fourteenth example. The wide occurrence and the remarkable structural similarity of the known  $\alpha/\beta$ -barrels raise in a new and acute form the fundamental question of whether proteins have similar folds because they are descended from a common ancestor or because of the structural requirements for stable folds.

In enzymes like TIM<sup>1</sup> and KDPG aldolase<sup>3</sup>, the subunit is formed by a single  $\alpha/\beta$ -barrel. The bifunctional enzyme *N*-(5'-phosphoribosyl) anthranilate iso-

merase-indole-3-glycerol-phosphate synthase has two  $\alpha/\beta$ -barrels with different catalytic activities in the same monomer unit<sup>4</sup>. In dimeric trimethylamine dehydrogenase<sup>5</sup> and tetrameric pyruvate kinase<sup>6</sup>, the  $\alpha/\beta$ -barrel forms one of the three domains that make up the subunit.

Although their structural environments and catalytic activities vary greatly, the  $\alpha/\beta$ -barrels for which the structural details are known are remarkably similar. The barrel  $\beta$ -sheet is always formed by eight parallel strands that slope at an angle of about 35° to the barrel axis. This slope means that adjacent strands are sheared relative to each other, as explained in Fig. 1. The extent of the shear is defined by the shear number, and for all  $\alpha/\beta$ -barrels this number is eight. The cross-section of the barrel varies in shape between circular, as in glycolate oxidase<sup>7</sup> where the barrel diameter is 14.5 Å, and elliptical, as in TIM<sup>1</sup> where the short and long diameters are 11.5 and 16.5 Å (see Fig. 1). Almost all the strands are linked by  $\alpha$ -helices which normally pack on the sheet with their axes parallel to the strand direction. (In some proteins, before the completion of the  $\alpha/\beta$ -barrel, the chain makes an excursion out of the barrel to form another domain or a large external loop.) In all cases, the active site is at the same end of the barrel in a cavity formed where the

carboxy termini of the strands loop back to the amino ends of the helices.

The  $\alpha/\beta$ -barrel fold is distinguished from other types of protein fold by the closed barrel structure formed by the hydrogen bonds in the parallel  $\beta$ -sheet. McLachlan analysed<sup>8</sup> the geometrical properties of an ideal  $\beta$ -sheet barrel structure and found that the number of the strands in the barrel, their slope to the barrel axis, the shear of the strands, the mean radius of the barrel and the mean

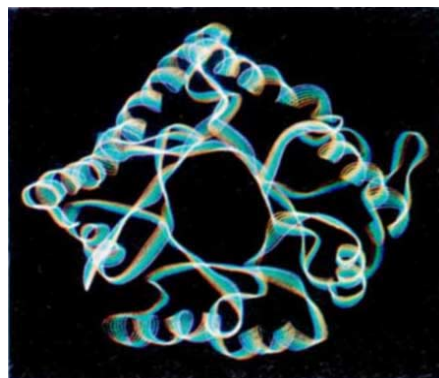


Fig. 2 The 8-fold  $\alpha/\beta$ -barrel of enolase, a common protein fold (courtesy of L. Lebioda).

twist of strands are related to each other by the equations in Fig. 1. Originally, McLachlan's work arose from an attempt to understand the structures formed by antiparallel  $\beta$ -sheets. However, antiparallel  $\beta$ -sheets form layer or sandwich structures, not barrels, and so the equations are not applicable to them<sup>10</sup>. Analysis of the protein structures that have since become available for parallel

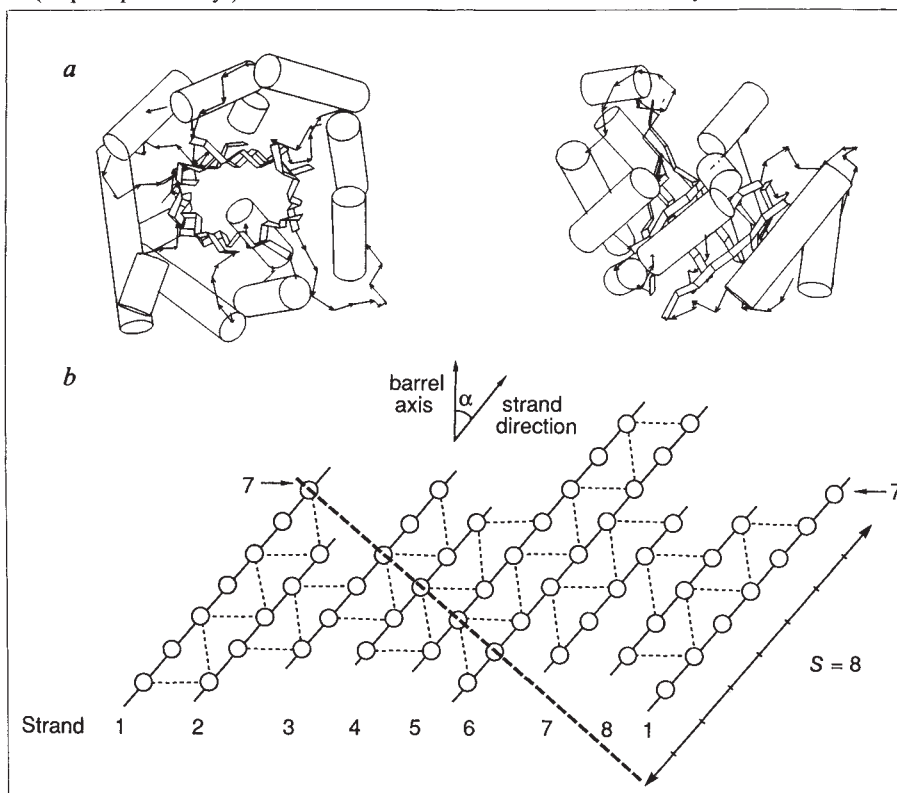


Fig. 1 Geometrical features of the  $\alpha/\beta$ -barrel structure. *a*, Two orthogonal views of the structure of triose phosphate isomerase<sup>1</sup>. Helices are represented by cylinders and strands of  $\beta$ -sheet by ribbons. *b*, Schematic drawing of the central  $\beta$ -sheet in TIM produced by unrolling the barrel. Circles, amino-acid residues; broken lines, hydrogen bonds. The first strand is shown twice to illustrate how the sheet closes upon itself to form the barrel.

The geometrical characteristics of the  $\alpha/\beta$ -barrel structures are:

$n$ , the number of  $\beta$ -sheet strands that form the barrel — 8;

$\alpha$ , the slope of the strands to the axis of the barrel — 35°;

$S$ , the shear number of the sheet. If the strands were vertical ( $\alpha = 0$ ), a line across the sheet, perpendicular to the strand direction, would come back on itself. Thus, if we started from residue 7 in strand 1 the line would return to the same residue. In the  $\beta$ -sheets of  $\alpha/\beta$ -barrels, such a line (see *b*) does not return to the same residue because of the slope of the strands. A line from a strand comes back to that strand at a position displaced by eight residues from where it started. This defines  $S$  as 8.

$R$ , Mean radius of the barrel. In  $\alpha/\beta$ -barrels this is close to 7.0 Å. For an ideal  $\beta$ -sheet barrel<sup>8</sup>, these quantities are related by the equations

$$\tan \alpha = 0.75S/n; R = 2.2/\sin(\pi/n) \cos(\alpha).$$

$\beta$ -sheets in  $\alpha/\beta$ -barrels shows that in these molecules McLachlan's equations do apply and that they can be used to help understand the principles governing such structures.

Knowledge of conserved features of  $\alpha/\beta$ -barrels helped in what is probably the most successful attempt so far to predict the three-dimensional structure of a protein from its amino-acid sequence. Crawford *et al.*<sup>9</sup> used standard algorithms to predict the secondary structures in ten homologous sequences of the  $\alpha$ -subunit of tryptophan synthase. They averaged the results for homologous positions and recognized the features of an  $\alpha/\beta$ -barrel. This, together with other data, allowed them to predict a model with the correct overall fold. The secondary-structure assignments for all nine helices and six of the eight  $\beta$ -strands are close to those which were later found in the crystal structure<sup>9</sup>.

Evolutionary relationships between proteins can be argued on the basis of the similarities in their sequences, structures and the geometries of their active sites. Most of the  $\alpha/\beta$ -barrels have little or no similarity in their amino-acid sequences, and so discussions of their evolutionary relationships have mainly involved questions of structure<sup>2</sup>. Proteins that have diverged from a common ancestor have structural differences that increase with the number of amino-acid substitutions<sup>11</sup>. The main elements of secondary structure and the active-site peptides retain their fold but may shift relative to each other. For two homologous proteins, the structural differences in these regions, expressed in terms of the root mean square (r.m.s.) difference in the position of main-chain atoms,  $\Delta$  (Å), are related to sequence differences by the approximate equation  $\Delta = 0.40e^{1.87H}$ , where  $H$  is the proportion of non-identical amino-acid residues<sup>11</sup>. Surface loops and peripheral elements of secondary structure may have quite different folds and no sequential homology.

Before the sequence of spinach glycolate was determined, Branden *et al.*<sup>7</sup> compared the structure of its  $\alpha/\beta$ -barrel with that of yeast flavocytochrome  $b_2$ . The two proteins use the coenzyme FMN to oxidize small  $\alpha$ -hydroxy acids. For the 320 residues that have the same fold, the r.m.s. difference in position is 1.6 Å. This implies<sup>7,11</sup> a sequence identity of 30–50 per cent. The recently determined sequence of glycolate oxidase<sup>12</sup> has 33 per cent of its residues identical to those in flavocytochrome  $b_2$ .

The two  $\alpha/\beta$ -barrels in *N*-(5'-phosphoribosyl) anthranilate isomerase-indole-3-glycerol-phosphate synthase catalyse two consecutive reactions in the biosynthesis of tryptophan. Both  $\alpha/\beta$ -barrels consist of about 200 residues. For the 185 residues that have the same fold in

## Enter the Chinese dragons

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

*Shunosaurus* is one of six magnificent dinosaurs from China on display at the British Museum (Natural History) in London from 17 June until 31 January 1989. By virtue of its unsurpassed Mesozoic record, China is "the most important dinosaur region in the world" says Angela Milner, who excavated in China in 1982 with other researchers from the museum. The six skeletons show a good cross-section of the riches of Chinese dinosaurs, from the Triassic prosauropod *Lufengosaurus* to the Upper Cretaceous hadrosaur *Tsintaosaurus*. Others include the 2.5-metre-tall carnosaur *Gasosaurus*, the stegosaur *Tuojiangosaurus* and the 30-tonne sauropod *Mamenchisaurus*, which had the longest neck of any animal known. The fossils have travelled 7,000 miles by courtesy of the Institute of Vertebrate Palaeontology and Palaeoanthropology in Beijing. A four-year joint research programme between the institute and palaeontologists in Canada is currently unearthing dinosaurs in the Gobi Desert, many of which are previously undiscovered species.

Henry Gee

the two domains, r.m.s. difference in position is 2.17 Å and the sequence identity is 10.3 per cent<sup>4</sup>. The close fit of these values to the equation implies the descent of the two domains from a common ancestor.

Normally, the use of the equation requires atomic coordinates that have been well refined with high-resolution data. Lebioda and Stec in their paper in this issue<sup>2</sup> use an approach that overcomes this requirement. An ideal  $\alpha/\beta$ -barrel has 8-fold symmetry (see Fig. 2), but real  $\alpha/\beta$ -barrels have large local deviations from ideal symmetry. Proteins descended from a common ancestor might be expected to share to some extent the same local deviations. Hence the proper superposition of their structures: first strand of one structure on the first strand of the other structure, and so on, should produce a better fit than improper superpositions: first strand on one structure on second strand of the other structure, and so on. Previously, Lebioda *et al.*<sup>13</sup> used such calculations to show that the  $\alpha/\beta$ -barrels of TIM, pyruvate kinase and KDPG aldolase are probably not descended from a common ancestor. In their new work<sup>2</sup>, Lebioda and Stec show that enolase and pyruvate kinase, which catalyse two consecutive reactions in the glycolytic

pathway, probably do share a common ancestor. Their conclusion is supported by similarities in the geometry of the active sites.

It is clear that some  $\alpha/\beta$ -barrels with very little similarity in amino-acid sequence are descended from a common ancestor. But the similarities in the structures of the other  $\alpha/\beta$ -barrels almost certainly arise from the stringent requirements of this fold — requirements that are not yet clearly understood. □

1. Banner, D.W. *et al.* *Nature* **255**, 609–614 (1975).
2. Lebioda, L. & Stec, B. *Nature* **333**, 683–686 (1988).
3. Mavridis, J.M., Hatada, M.H., Tulinsky, A. & Lebioda, L. *J. molec. Biol.* **162**, 419–444 (1982).
4. Priestle, J.P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 5690–5694 (1987).
5. Lim, L.W. *et al.* *J. biol. Chem.* **261**, 15140–15146 (1986).
6. Stuart, D.I., Levick, M., Muirhead, H. & Stammers, D.K. *J. molec. Biol.* **134**, 109–142 (1979).
7. Branden, C.-I. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **52**, 491–498 (1988).
8. McLachlan, A.D. *J. molec. Biol.* **128**, 49–79 (1979).
9. Crawford, J.P., Niermann, T. & Kirschner, K. *Proteins* **2**, 118–129 (1988).
10. Chothia, C. & Janin, J. *Biochemistry* **21**, 3955–3965 (1982).
11. Chothia, C. & Lesk, A.M. *EMBO J.* **5**, 823–826 (1986).
12. Cederlund, E., Lindqvist, Y., Soderlund, G., Branden, C.-I. & Jornvall, H. *Eur. J. Biochem.* **173**, 523–530 (1988).
13. Lebioda, L., Hatada, M.H., Tulinsky, A. & Mavridis, J.M. *J. molec. Biol.* **162**, 445–458 (1982).

Cyrus Chothia is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH and the Department of Chemistry, University College London, Gordon Street, London WC1 0AJ, UK.



## Solid-state physics

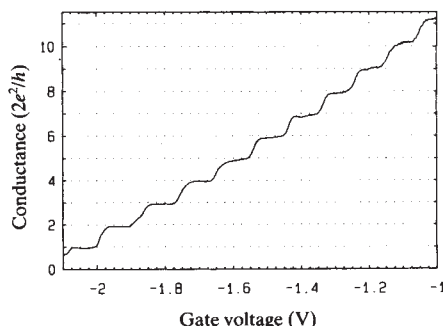
## Quantized ballistic resistance

F.W. Sheard and L. Eaves

QUANTUM interference effects are of great interest because they are a direct manifestation of the wave properties of matter. Such effects are not easily observed using conduction electrons in solids, because they are degraded by dissipative scattering (resistance) which destroys the coherence of the wavefunction. But advances in crystal growth and processing techniques now permit the fabrication of small-scale, high-quality semiconductor devices in which coherence effects are clearly revealed. These devices generally consist of multiple thin layers of different semiconductor material (heterostructures), and at low temperatures (around 4 K) electrons can travel ballistically along paths of about 1,000 Å, which are mesoscopic<sup>1</sup> (intermediate between atomic and macroscopic size). Quantum effects can then become important when the quantum-mechanical de Broglie wavelength associated with the ballistic motion becomes comparable with some physical dimension (layer thickness or width) of the heterostructure<sup>2,3</sup>. Two groups, van Wees *et al.*<sup>4</sup> and Wharam *et al.*<sup>5</sup>, now report that when the ballistic flow of electrons is restricted by a narrow channel, the electrical resistance is quantized and takes values  $R_p = h/2e^2p$  ( $p = 1, 2, 3 \dots$ ) which depend only on the fundamental constants  $h$  (Planck's constant) and  $e$  (the charge on an electron).

The pursuit by semiconductor manufacturers of devices with very fast switching speeds has led to the study of transistor

heterostructures based on layers of GaAs and (AlGa)As, in which the electron transport is perpendicular to the layers. The switching speed is increased by reducing the transit time of an electron through the base region, which depends on the base width and carrier velocity. In a heterostructure, the base is a very thin GaAs



**Fig. 2** Conductance  $G$  as a function of gate voltage (after subtraction of lead resistance), showing plateaux at multiples of  $2e^2/h$ . (From ref. 4; similar results are reported in ref. 5.)

layer and fast ballistic, rather than slow diffusive, transport can be attained.

Experiments have demonstrated that the transport<sup>1</sup> through GaAs base regions is ballistic or collisionless for thicknesses up to 800 Å. The thin GaAs layer is sandwiched between two (AlGa)As layers and behaves as a potential well to the electrons. Quantum coherence of the wavefunction across this layer then gives rise to standing-wave states at certain quantized values of the electron energy. These states are analogous to the standing-wave modes of vibration of a finite length of stretched string and can be regarded as arising from interference between the electron waves reflected back and forth between the boundaries of the layer.

Electrons are injected by quantum tunnelling through an (AlGa)As layer, which allows the injection energy to be varied by changing the applied voltage. Resonant peaks in the current then occur when the energy of the injected electron matches that of a standing-wave quantum state<sup>2</sup>. Experiments using *p*-type rather than *n*-type GaAs have recently shown similar resonances<sup>3</sup> resulting from injection of light holes (electron vacancies) into the quantum states of a GaAs layer.

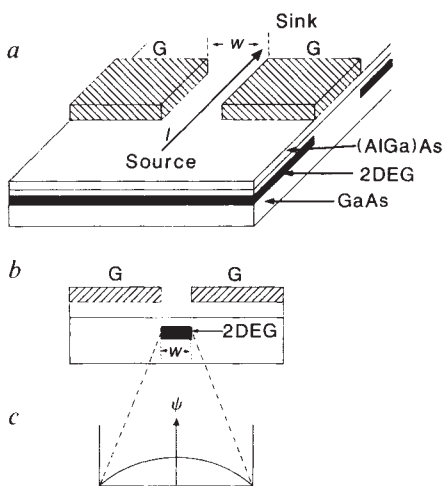
Because coherence effects are destroyed by dissipative scattering it is surprising to find that dissipation itself can be quantized. van Wees *et al.*<sup>4</sup> and Wharam *et al.*<sup>5</sup> found this recently in some remarkable experiments on current flow through a narrow channel in which the conductance  $G = 1/R$  is quantized in units of  $2e^2/h$ .

In these experiments (Fig. 1) the current flow is parallel to a GaAs-(AlGa)As interface in a suitably doped heterostructure and takes place in a thin layer (about 100 Å thick) in which the electron motion is essentially two-dimensional. The current flow along this two-dimensional electron gas is directed to pass through a narrow channel by means of gate electrodes evaporated on top of the heterostructure<sup>6</sup>.

Altering the gate voltage removes the two-dimensional electron gas below the electrodes and also varies the effective width  $w$  of the channel (Fig. 1b). As  $w$  is varied, the conductance  $G$  exhibits plateaux at multiples of  $2e^2/h$  (Fig. 2). When  $w$  is sufficiently small, so that the electron motion through the channel is essentially one-dimensional, the conductance takes the fundamental value  $G_1 = 1/R_1 = 2e^2/h$ . The quantized values of resistance  $R_p = h/2e^2p$  ( $p = 1, 2, 3 \dots$ ) are the same as those found in the quantum Hall effect<sup>7</sup>, which occurs when a high magnetic field is applied perpendicular to the plane of a two-dimensional electron gas. The quantum Hall effect, however, is a dissipationless phenomenon. Although the Hall resistance, defined as the ratio of the Hall voltage to the current flow, is similarly quantized, the current flows without dissipation in a direction perpendicular to the Hall potential drop.

In the new experiments<sup>4,5</sup> on the quantization of dissipative resistance, it is necessary that the electron flow through the channel is ballistic and is therefore described by a coherent wavefunction. Earlier experiments in which the flow was diffusive did not show the quantization effect. The dissipation of energy must occur via inelastic processes in thermalizing reservoirs which serve as the source and sink of the current flowing through the channel (see Fig. 1). The ballistic flow of electrons between the two reservoirs<sup>8,9</sup> at different voltages is analogous to the transfer of radiation between two black bodies at different temperatures. The temperature difference corresponds to the voltage difference  $V$  between the reservoirs, which raises the energy of electrons in the source reservoir relative to those in the sink by an amount  $e \times V$ .

Heat transfer from a hotter to a cooler black body occurs essentially because the hotter body radiates more light than the cooler. Similarly, more electrons travel ballistically from the source at the higher energy than from the sink. The imbalance gives rise to a current  $I$  which is proportional to the energy difference  $eV$ . Using Fermi-Dirac statistics, it can be shown that for a strictly one-dimensional channel at low temperatures,  $I = 2e^2V/h$ . The second power of  $e$  arises from the quantum of charge carried by each electron and the factor 2 from the two spin states per electron. The appearance of Planck's constant  $h$  is a consequence of the kinematics



**Fig. 1** *a*, Geometry of heterostructure showing two-dimensional electron gas (2DEG) below (AlGa)As/GaAs interface, direction of current flow  $I$  in the 2DEG and gate electrodes (G). *b*, Cross-section showing depletion of 2DEG by negative potential applied to gates and formation of narrow channel of width  $w$ . *c*, Lowest-energy standing-wave state  $\psi$  describing lateral confinement of electrons in channel.

of a one-dimensional quantum channel. So the fundamental resistance  $R_1 = V/I = h/2e^2$ , the value obtained experimentally.

The flow through the channel is one-dimensional when all the electrons are described by the lowest-energy transverse standing-wave state appropriate to the width  $w$  of the channel (Fig. 1c). When  $w$  is increased, the energies of the transverse states fall and more such states (the higher excitations) are occupied by electrons. Because each transverse state gives the same conductance,  $G_1 = 2e^2/h$ , and the contributions add in parallel, this accounts for the equally spaced conductance steps shown in Fig. 2. The quantization of dissipation is thus a novel consequence of the quantum coherence of the wavefunction.

It is interesting that quantum steps in resistance have also been reported when

breakdown of the dissipationless plateaus of the quantum Hall effect occurs<sup>10</sup>. The magnitude of these steps is a small fraction of  $h/2e^2$  and seems to be related to changes in quantum number of electrons during dissipative transitions<sup>11</sup>, but a complete theory has not yet been derived. □

1. Heiblum, M. *et al.* *Phys. Rev. Lett.* **55**, 2200–2203 (1985).
2. Heiblum, M. *et al.* *Phys. Rev. Lett.* **58**, 816–819 (1987).
3. Heiblum, M. *et al.* *Phys. Rev. Lett.* **60**, 828–831 (1988).
4. van Wees, B.J. *et al.* *Phys. Rev. Lett.* **60**, 848–850 (1988).
5. Wharam, D.A. *et al.* *J. Phys. C21*, L209–L214 (1988).
6. Thornton, T.J. *et al.* *Phys. Rev. Lett.* **56**, 1198–1201 (1986).
7. von Klitzing, K., Dorda, G. & Pepper, M. *Phys. Rev. Lett.* **45**, 494–497 (1980).
8. Büttiker, M. *et al.* *Phys. Rev. B31*, 6207–6215 (1985).
9. Landauer, R. *Phil. Mag.* **21**, 863–867 (1970).
10. Bliok, L. *et al.* *Surf. Sci.* **196**, 156–164 (1988).
11. Eaves, L. & Sheard, F.W. *Semicond. Sci. Technol.* **1**, 346–349 (1986).

F.W. Sheard and L. Eaves are in the Department of Physics, University of Nottingham, Nottingham NG7 2RD, UK.

## Quantum optics

# Squeezed and non-classical light

R.K. Bullough

THE development of lasers with a controllable phase operating at powers where materials respond in a nonlinear way makes possible studies of light which displays strictly non-classical features. Under these circumstances, Maxwell's classical equations of electromagnetism no longer hold, and quantum theories, which are intrinsically 'noisy', have to be invoked. At two recent meetings\*, laser physicists discussed ways of reducing the noise levels in photodetection by 'squeezing' the light, so improving the sensitivity of optical experiments and the quality of optical-signal transmissions.

In Dirac's quantum theory of light, the various vibrational modes of the electromagnetic field, conveniently labelled by their wavenumber  $k$  ( $2\pi$  times the reciprocal of their wavelength  $\lambda$ ), behave like harmonic oscillators. The energy contained by such oscillators can have only certain allowed values,  $(n + \frac{1}{2})\hbar\omega_k$ , where  $n=0,1,2,\dots$ ,  $\hbar$  is Planck's constant divided by  $2\pi$ , and  $\omega_k = ck$  is the frequency of the oscillator ( $c$  is the speed of light). According to Dirac, this is equivalent to saying that the mode contains  $n$  photons, each of energy  $\hbar\omega_k$ . The state of the mode, an eigenstate because it is the exact solution of the quantum mechanical equations giving the number  $n$ , is labelled  $|n\rangle$  and is sometimes termed a Fock state.

Even when a mode contains no photons, and is in a state described as the vacuum state  $|0\rangle$ , it still has zero-point energy  $\frac{1}{2}\hbar\omega_k$ . This can be understood in terms of Heisenberg's uncertainty principle. The two components of the electro-

magnetic field in the mode, the electric field  $E_k$  and the magnetic vector potential  $A_k$ , have an average value zero. But their fluctuations  $\Delta E_k$  and  $\Delta A_k$  do not vanish, and according to the uncertainty principle,  $\Delta E_k \Delta A_k \geq \frac{1}{2}\hbar$ . The vacuum does have the minimum uncertainty, however, for  $\Delta E_k \Delta A_k = \frac{1}{2}\hbar$ . Moreover, both  $\Delta E_k$  and  $\Delta A_k$  have the value  $(\frac{1}{2}\hbar)^{1/2}$ .

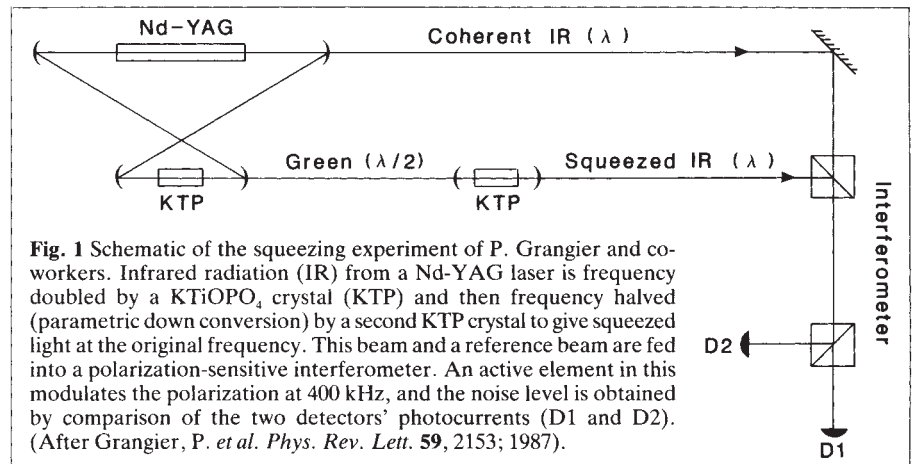
The most classical non-vacuum states, the coherent or  $\alpha$ -states  $|\alpha_k\rangle$ , have the same fluctuations as the vacuum and are also minimum uncertainty states. A single-mode laser well above threshold is in such

photons — in contrast with the normal 'bunching' of photons observed in black-body radiation.

The production of particular non-classical fields is important for reducing the uncontrolled variations in field intensities to improve the sensitivity of optical experiments. Fluctuations in vacuum or excited-field states are observed as quantum noise (shot noise) in the photo-detection current measuring the electric field, and this seems to place a limit on the precision of intensity measurements. If the electric field could be put in an eigenstate, its fluctuation  $\Delta E_k$  would vanish and it would have zero quantum noise. Although no such states have been made, there are circumstances in which  $\Delta E_k < (\frac{1}{2}\hbar)^{1/2}$ , and such light is 'squeezed'. This is achieved at the expense of the other field quadrature  $A_k$ , so that  $\Delta A_k > (\frac{1}{2}\hbar)^{1/2}$  to satisfy the uncertainty principle.

J.F. Kimble (Univ. Texas, Austin) uses lithium niobate crystals with nonlinear optical properties to split single photons into two photons with half the energy (see the News and Views article by D.F. Walls in *Nature* **324**, 210; 1986). Performed in an optical cavity, such 'parametric down conversion' currently reduces observed noise levels by 63 per cent. Kimble has now incorporated this technique in precision interferometry, as has P. Grangier of AT&T Bell Laboratories who uses a neodymium-YAG laser and a KTiOPO<sub>4</sub> crystal (Figs 1 and 2).

Grangier also noted the use of squeezed states to enhance the optical interferometry used in tests of general relativity. Gravitational waves in the Galaxy should cause the two arms of an interferometer to vibrate. G. Leuchs (Max Planck Institut



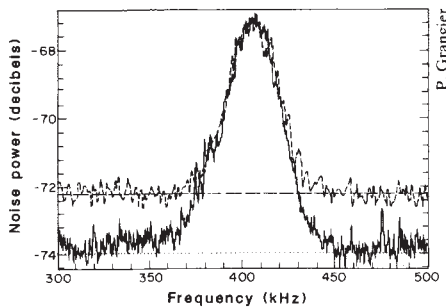
**Fig. 1** Schematic of the squeezing experiment of P. Grangier and co-workers. Infrared radiation (IR) from a Nd-YAG laser is frequency doubled by a KTiOPO<sub>4</sub> crystal (KTP) and then frequency halved (parametric down conversion) by a second KTP crystal to give squeezed light at the original frequency. This beam and a reference beam are fed into a polarization-sensitive interferometer. An active element in this modulates the polarization at 400 kHz, and the noise level is obtained by comparison of the two detectors' photocurrents (D1 and D2). (After Grangier, P. *et al.* *Phys. Rev. Lett.* **59**, 2153; 1987).

a state: it contains many photons although their number is indefinite. The probability of observing  $n$  photons follows a Poisson distribution: in particular the mean number  $\bar{n}_k (= |\alpha_k|^2)$ , the square of the field amplitude) equals the variance. Fields can now be made in the laboratory which are strictly non-classical in that their variance is less than their mean. These sub-poissonian fields contain antibunched

für Quantenoptik (MPI)) showed that when the sensitivity of his interferometer  $\delta l/l$  ( $l$  is the length of the interferometer arms) is increased from its present value of  $3 \times 10^{-18}$  to  $3 \times 10^{-22}$ , it should be able to respond to the 10–15 supernovae optimistically estimated to occur in our Galaxy each year. By modifying the shot-noise limit using the squeezing achieved so far, the sensitivity could be further improved

\* Squeezed and non-classical light 25–29 January 1988, and Photon localisation, detection, amplification and antibunching 21–22 January 1988, Cortina d'Ampezzo, Italy.





**Fig. 2** Noise reduction in the experiment (Fig. 1) of Grangier *et al.*, observed as the change in signal intensity close to the polarization-modulation frequency 400 kHz. Signal to noise  $S/N$  is increased from 5 dB ( $S/N = 3.16$ , upper trace) to 7 dB ( $S/N = 5.01$ , lower trace) when the squeezed light is switched on.

by a factor of about 3 — important because the distance seen into the Galaxy gains a factor of 3 and the volume sampled (determining the detection rate)  $3^3 = 27$ .

Microwave radiation from a black-body source at 4.2 K has been squeezed by 42 per cent, corresponding to a drop of 1.8 K below the 4.2-K noise floor (B. Yurke, AT&T). The device used was a parametric amplifier similar in principle to Kimble's, but using a superconducting quantum interference device (SQUID) as the nonlinear feature. Pulsed squeezed light could be useful for optical information transfer. The quantum noise that typically accompanies light-pulse signals in optical fibres can now be reduced, improving the signal-to-noise ratio and so clarifying the signal (R.E. Slusher, AT&T).

The squeezed vacuum, a squeezed state in which the average  $\bar{E}_0$  of the electric field is zero, also has impressive properties. For example, an excited atom spontaneously emits radiation into the many modes of the squeezed vacuum and suffers a small shift of its energy levels. This corresponds to the radiative shift in the ordinary vacuum first observed in hydrogen by Lamb and Retherford (*Phys. Rev.* **72**, 241; 1947). In hydrogen, the spherically symmetric atomic  $2s$ -state is shifted from the  $2p$ -states by about 1,000 MHz by interaction with the ordinary vacuum. This value would be changed in the squeezed vacuum. The rate of spontaneous emission is also modified, by a phase factor  $e^{i\phi}$ , determined by the parameters of the squeezed vacuum; by choosing the phase  $\phi = \pi/2$ , spontaneous emission can be much reduced. Likewise, peak widths in the spectrum of resonance fluorescence can be modified (narrowed) in a phase-dependent way (Carmichael, H.J. *et al. Phys. Rev. Lett.* **58**, 2539–2542; 1987).

H. Walther (MPI) reported the observation of quantum revivals in the single  $^{85}\text{Rb}$ -atom maser. This  $10^{-20}$  W maser passes a beam of excited  $^{85}\text{Rb}$  atoms through a superconducting niobium cavity with almost perfectly reflecting walls at 0.5 K. The atoms enter in a high Rydberg state (the  $63p$ -state in which the optical

electron orbits at a large radius of about 210 nm) and the cavity is tuned so that each atom can make a microwave transition at a wavelength  $\lambda = 1.38$  cm. The high quality-factor of the cavity means there is little damping and the low temperature means there are few black-body photons (0.15 on average). The photons emitted by the atoms build up a field which can return its photons to the atoms.

The system therefore oscillates (as a maser) but the revivals (return of the atoms to their excited  $63p$ -states) do not make a simple oscillation and are fundamentally quantum mechanical. Moreover the field in the cavity differs from that of the single-mode optical laser: it can be sub-poissonian, or it can involve atoms leaving the cavity in their upper states ( $63p$ ) after emitting and reabsorbing a photon once, twice, . . . . The probability distribution for the number of photons in the cavity should sharpen as more atoms pass through, eventually resembling the statistics of a Fock state  $|n\rangle$  with  $n \approx 60$ .

Walther also reported antibunching of radiation and the associated sub-poissonian statistics in the resonance fluorescence from a single  $^{24}\text{Mg}^+$  ion making  $\lambda = 0.28$   $\mu\text{m}$  transitions in a radio-frequency Paul ion trap. Equally remarkable is the observation of phase transitions to a 'Wigner' crystal by 2–50 laser-cooled  $\text{Mg}^+$  ions ordering on cooling (see *Nature* **331**, 208; 1988). Squeezing, predicted for both the one- and several-atom maser, is still to be observed, perhaps by an adaptation of Yurke's SQUID microwave methods.

There are coherent states other than the  $\alpha$ -states which are squeezed (unlike the  $\alpha$ -state). One can be generated using a parametric amplifier. But although non-linearity seems to be essential for squeezing light, R. J. Glauber (Harvard Univ.) showed that even linear dielectrics can reveal some paradoxes. It is well known that a dielectric has a new vacuum with frequencies different from those of the free unsqueezed vacuum (the resulting change in the zero-point energies explains the binding energy of atoms in the dielectric). Glauber showed that the vacuum of the dielectric is derived from the free vacuum by a squeezing transformation and so contains photons of the free vacuum. Is the vacuum squeezed, and can its photons be detected? E.R. Pike (Royal Signals and Radar Est.) discussed how photon localization could be investigated. The detection of one of a pair of correlated photons could help to determine the localization of the other. But does this raise the spectre of Einstein's EPR paradox (Einstein, A., Podolsky, D. & Rosen, N. *Phys. Rev.* **47**, 777–780; 1935) and the problem of action at a distance?  $\square$

R.K. Bullough is in the Department of Mathematics, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD, UK.

## Daedalus

# Internal sobriety

OUR blood carries many kinds of cell in suspension, and will even tolerate artificial ones. Small 'microspheres' of latex, albumin, polystyrene and so on can be safely injected into the circulation. Some have been coated experimentally with specific drugs, aimed at organs which delay the microspheres in their travels.

Daedalus is now coating such microspheres with enzymes: the alcohol and aldehyde dehydrogenases which oxidize alcohol in the liver. His idea is to counter alcoholism. Suspensions of microspheres bearing the immobilized enzymes will be injected into DREADCO volunteers, who will then try to get drunk. If all goes well, they will fail dismally. As fast as alcohol enters their bloodstream, it will be burnt up by the enzymatic microspheres; the most enthusiastic drinking should fail to shake their stern sobriety. It will, however, warm them up a bit. A standard drink has about 10 grams of alcohol in it. Fed to the vigilant blood-microspheres over (say) 20 minutes, it would increase the body's heat-output by about 240 watts during that period — trebling the normal metabolic rate.

As an instant sobering-up injection, Daedalus's microsphere suspension will be even more drastic. A drunk driver at the legal limit would not only be hit by about 500 watts for several minutes as the microspheres burnt the alcohol from his blood — he would collapse gasping in near-suffocation as they used up all the oxygen in his bloodstream to do it. Criminologists and aversion therapists who espouse the 'short sharp shock' theory of punishment, should applaud the new treatment.

With proper sizing and composition, enzymatic microspheres should degrade in the blood very slowly. Their coating of normal body enzymes should not arouse the suspicions of the body's immunological defences, and they should be far too big to diffuse into the waste-disposal cells of the liver. They might well enforce sobriety for days or weeks on end.

DREADCO's biochemists hope to extend their therapeutic range, and are coating suitable microspheres with enzymes to oxidize nicotine, cannabinal or cocaine, so as to reclaim the wretches now in thrall to tobacco, cannabis and 'hard' drugs.

But even Daedalus has doubts about coating his microspheres with the enzymes that oxidize glucose, and injecting them as a 'slimming-shot'. Glucose is the general fuel of the body. Burning it pre-emptively in the blood could effortlessly cancel the most obsessive overeating. But the resulting wasting fever might simply replace obesity by a form of galloping, breathless *anorexia nervosa*. David Jones

## Syphilis in a Pleistocene bear?

SIR—Rothschild and Turnbull<sup>1</sup> attribute the cause of exuberant bone formation and periosteal reaction in the humerus, ulnae and three thoracic vertebrae of a Pleistocene bear (*Arcdotus simus*) to a treponemal infection. They use pathological bone morphology and immunofluorescent analysis as a basis for this conclusion. There is another view, that these lesions were caused by a mycotic (fungal) infection, tuberculosis or an unknown pathogen<sup>2</sup>.

The spinal lesions involve vertebrae T3, T4 and T5. Lytic destruction of the anterior half of each vertebral body has left a rough surface which is poorly defined radiographically (see figure). A large mass of hyperplastic bone circles the ventral portion of each vertebra and loosely locks the vertebrae together. The intervertebral disks appear to be destroyed. The neural canals, crests, spines, transverse processes and ribs are not affected. I observed no sequestra, vertebral collapse or rib lesions. There are several round, 'punched-out' lesions on the mandible which are 3 mm in diameter. The distal third of each ulna shows 1-cm diameter ovoid lesions penetrating 2 cm into the bone. There is minor periosteitis and a fistulous tract on the right ulna. Also noted is some periosteal reaction near the head of one humerus.

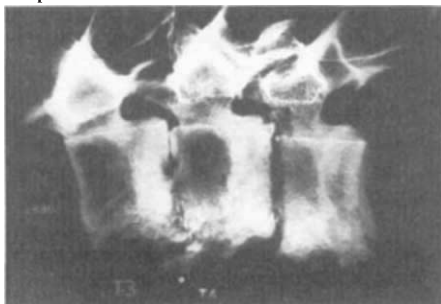
Treponematoses has been reported in man and monkey<sup>3</sup>. To my knowledge, this is the first and only report suggesting infection in modern or prehistoric subprimates. The description of the lesions are atypical for treponemal infections which favour involvement in multi-joints, the skull and tibiae<sup>4</sup>. Five per cent of osseous treponematoses occurs in the spine where proliferation of bone is limited and affects all adjoining structures<sup>4</sup>. Fusion and direct lesion extension is not seen in this specimen. The vertebrae are not fused and exhibit anterior body lysis. There is notable non-involvement of the ribs, neural arches and transverse processes — often seen in tuberculosis<sup>4,5</sup>.

Actinomycosis, blastomycosis and coccidioidomycosis may produce these lesions singly or in combination<sup>4,5</sup>. They may affect the animal locally or systemically. These mycotic agents are endemic throughout the mid-west United States and often infect domestic and wild animals<sup>6</sup>.

The immunofluorescent-histological analysis used indicates treponema; however, the test's reliability is questionable because: (1) the specimen is over 11,000 years old and treponema antigen is unlikely to maintain its reactivity; (2) the specimen was partially fossilized, leaving no fresh bone or soft tissue further reducing the chance of antigen survival; and (3) having been exposed to wildlife, then

buried for thousands of years and handled by innumerable people, there is a high probability of contamination with a cross-reactive agent. It would be helpful if samples from other areas of the *Arcdotus* skeleton, as well as from other skeletons of a similar geological age and geography were histologically analysed and compared.

It is unlikely that this is a case of treponematoses. No authenticated tre-



Radiograph of T3, T4, T5 showing anterior vertebral body destruction; a condition seen in tuberculosis and some fungal infections but absent in treponematoses.

ponemal infections have been found in a non-primate. The lesions in this prehistoric bear have similar signs to those made by fungal or tubercular disease; either as a single or multiple infection<sup>4,5</sup>. The immunofluorescent analysis is limited to one area of the specimen and depends on the antigen remaining reactive and the bone uncontaminated over a 11,000-year span. Infective organisms tend to change in virulence and mode of action with time. This specimen may have been infected with unknown organisms or an early precursor of modern pathogens which resulted in a different expression of bone repair as compared with modern animals.

E.J. NEIBURGER

Curator-Lake County Museum,  
1000 North Avenue,  
Waukegan, Illinois 60085, USA

1. Rothschild, B. M. & Turnbull, W. *Nature* **329**, 61–62 (1987).
2. Neiburger, E. J. *Paleopathol. Newsletter* **48**, 8–11 (1984).
3. Hackett, C. J. *Bull. WHO* **29**, 7–14 (1963).
4. Steinbock, R. T. *Paleopathologic Diagnosis and Interpretation*. 89–169 (Thomas, Springfield, 1976).
5. Ortner, D. & W. Putschar. *Identification of Pathological Conditions In Human Skeletal Remains* 180–227 (Smithsonian Institution, Washington DC, 1981).
6. West, G. P. (ed.) *Encyclopedia of Animal Care* 11th edn (Williams and Wilkins, London, 1975).

## Right-handed longevity not so sinister after all?

SIR—Smoking, eating saturated fats, stress . . . and now we are told that left-handedness curtails the life-span. As an incurable sinistral, I suppose that I have a vested interest in finding Halpern and Coren's answer to the question "Do right-handers live longer?" (*Nature* **333**, 213; 1988) unconvincing. However, it is well established that nowadays there is considerably less pressure on children to use their right hands than there used to be.

Any 80-year old will tell you that he or she was not given any choice as to which hand to use for writing; I am sure that this was equally true in the teaching of baseball technique. Older baseball players probably have a greater tendency to use their right hands, through 'nurture' rather than 'nature'. This is surely more than adequate as an explanation of the mere eight months difference in the mean age at death of right- and left-handed baseball players, especially in view of the vast standard deviations.

RICHARD J. PERRY

Physiological Laboratory,  
University of Cambridge,  
Cambridge CB2 3EG, UK

SIR—The reason why Porac and Coren, mentioned in the letter from Halpern and Coren (*Nature* **333**, 213; 1988), found fewer left-handers in their 50s and none at the age of 80 or above might be very simple: in former times left-handers were forced to become right-handers — at least in Austria and Germany.

In my youth in Vienna — I was born in 1923 — I never saw anyone writing with

the left hand. I was astonished, in the 1950s, to observe some Americans writing obliquely with their left hand. Obviously, parents and teachers in the United States, on the advice of psychologists, let left-handed children have their way earlier than in Central Europe. Even now, left-handed writing is a rare exception in Austria and found only in the younger generation. Genetic or racial differences cannot be the reason for this difference between Austria and America.

I think that the conclusions of Halpern and Coren are based on wrong assumptions. The "higher environmental risk" they assume existed was only the strong hand of father and mother who did not want their child to be 'abnormal'.

FRIEDRICH KATSCHER

Mariahilfer Strasse 133,  
A-1150 Vienna, Austria

SIR—Perhaps the strenuous movements involved in baseball batting and pitching have graver physiological sequelae for left-handers than right-handers. There are other biases that should make one wary of extrapolating the conclusion of Halpern and Coren (*Nature* **333**, 213; 1988) to other groups of people. Their subjects were presumably men and very likely most of them were raised in North America. We are left with the inference that left-handed women cricketers need not worry unduly.

W. W. NICHOLS

Public Health Laboratory,  
John Radcliffe Hospital,  
Oxford, OX3 9DU, UK



## Pristane/phytane ratio as environmental indicator

SIR—The discussion by ten Haven *et al.*<sup>1</sup> concerning the use of pristane (Pr) to phytane (Ph) ratio as an environmental indicator has missed the point of much of the early literature, and misconstrues the significance of some recent data. The problem of error in the measurement of Pr concentrations caused by coelution of 2,6,10-trimethyl-7-(3-methylbutyl)-dodecane must be kept in perspective. The compound is poorly resolved from Pr on most good-quality capillary columns and has been identified in only one very unusual and geologically young crude oil and in few recent sediments<sup>2</sup>. There is not sufficient evidence that it is present in sufficient concentration to influence significantly the measurement of Pr concentrations in most ancient sediments and oils.

Bituminous coals and oils originating from higher plants are known<sup>3,4</sup> to have high Pr/Ph ratios (5–10) in comparison with marine oils and sediments (<1–3). Further, Pr/Ph ratios increase from 1–2 in brown coals to a maximum of 4–10 in bituminous coals, and significant Pr generation occurs at the threshold of oil generation. These observations were rationalized in terms of the preferential formation of phytanic acid by oxidation of phytol, followed by the conversion of phytanic acid to Pr during catagenesis<sup>3,4</sup>, and have been completely ignored by ten Haven *et al.*<sup>1</sup>. Clearly, high Pr/Ph ratios ( $\geq 3$ ) are indicative of an input from terrestrial organic matter that has a high probability of being exposed to oxidation before or during deposition. But as the depositional conditions and source inputs control the distribution of Pr and Ph precursors, it is inappropriate for Pr/Ph ratios to be used as a redox indicator in low-maturity samples without consideration of the precursors and the maturity level.

Most marine, organic-rich sediments and oils have a narrow range of Pr/Ph ratios (0.8–2.5)<sup>5</sup>, and with increasing maturity there can be a gradual increase in Pr/Ph ratio within this range. This has been associated in one case<sup>6</sup> with other evidence of oxidation of organic matter during deposition. The proposal that Pr/Ph ratios cannot reflect the redox conditions of the final resting place of organic matter<sup>1</sup> is a question of semantics because all sediments ultimately become reducing during burial; it is the depositional history that is the critical factor in determining the extent of oxidation of organic matter, not the final resting place.

Pr/Ph ratio cannot be used as a palaeo-environmental indicator at low maturity level<sup>2</sup> and this is implicit in the early literature<sup>3,4</sup>. At higher maturation levels, high Pr/Ph ratios ( $> 3.0$ ) are reliable indicators of input of terrestrial organic matter<sup>4</sup> and very low values are character-

istic of anoxic, and often hypersaline, environments<sup>1,2</sup>. The significance of values between 0.8 and 2.5 in mature normal marine sediments and oils is more difficult to assess. Whether the origin of pristane is ultimately phytol or tocopherols<sup>7</sup> may be irrelevant to the application of the Pr/Ph ratio as a redox indicator. Tocopherols are claimed to have greater preservation potential than phytol and its derivatives<sup>7</sup>. The effect of some oxidative degradation of organic matter during deposition would be selectively to preserve tocopherols, giving rise to higher Pr/Ph ratios during subsequent maturation. The use of Pr/Ph ratios as a correlative tool was originally based on many empirical observations which have now expanded greatly and are still valid. Any qualification to the explanations for the origins of Pr and Ph do not invalidate the fundamental observations.

T.G. POWELL

*Division of Continental Geology,  
Bureau of Mineral Resources Geology  
and Geophysics,*

*GPO Box 378, Canberra ACT, Australia*

TEN HAVEN *ET AL.* REPLY — We agree with Powell that pristane (Pr) to phytane (Ph) ratios close to unity should be interpreted with great caution and that fewer difficulties arise with extremely high (as in coals) or extremely low (as in sediment from hypersaline environments) Pr/Ph ratios. In these latter cases, the Pr/Ph ratios are probably influenced by specific sources of the organic matter and specific depositional environments, where the redox condition is only one of the factors.

The main point of our paper<sup>1</sup> concerns the generalizations and oversimplifications implied in the use of the Pr/Ph ratio as an indicator of the level of oxygen at the site of organic matter deposition into sediments. In much recent work (mainly that dealing with sediments containing immature organic matter), a Pr/Ph ratio of unity has become a kind of magic number to differentiate between palaeo-oxygen levels, commonly with reference to Didyk *et al.*<sup>8</sup>. We made several new arguments<sup>1</sup> based on novel geochemical findings to emphasize the restricted use of the Pr/Ph ratio as a palaeoenvironmental indicator. Note that the intermediates in the diagenetic scheme of Didyk *et al.*<sup>8</sup>, who studies the conversion of phytol in a few recent sediments, have not been verified in many cases. Moreover, it is becoming obvious that many precursors of both pristane and phytane have to be taken into account<sup>2,7</sup>. The selective preservation of these precursors will be different in different sedimentary environments. In other words, the ultimate Pr/Ph ratio is determined by a large variety of known and unknown precursors, their selective preservation and their pathways of diagenesis.

The coelution of pristane with 2,6,10-trimethyl-7-(3-methylbutyl)-dodecane on apolar stationary phase is a new result.

Sediments containing mature organic matter and a few crude oils have been checked for the presence of this compound. In addition to other examples<sup>2,9</sup> sapropels from an open marine environment in the Mediterranean (typical potential oil-source rocks) are known to contain this new compound in concentration three times as high as pristane<sup>10</sup>.

Of course, we do not object to using the Pr/Ph ratio as a correlative tool when samples of the same origin and maturation level are compared with one another (as in oil/source-rock correlation).

H.L. TEN HAVEN

J. RULLKÖTTER

*Institute of Petroleum and Organic*

*Geochemistry (ICH-5),*

*KFA Jülich,*

*P.O. Box 1913, D-5170 Jülich, FRG*

J.W. DE LEEUW

J.S. SINNINGHE DAMSTÉ

*Department of Chemistry and*

*Chemical Engineering,*

*Delft University of Technology,*

*de Vries van Heystplantsoen 2,*

*NL-2628 RZ Delft, The Netherlands*

1. ten Haven, H.L., de Leeuw, J.W., Rullkötter, J. & Sinninghe Damsté, J.S. *Nature* **330**, 641–643 (1987).
2. Volkman J.K. & Maxwell, J.R. in *Biological Markers in the Sedimentary Record* (ed. Johns, R.B.) 1–42 (Elsevier, Amsterdam, 1986).
3. Brooks, J.D. *et al.* *Nature* **222**, 257–259 (1969).
4. Powell, T.G. & McKirdy, D.M. *Nature* **243**, 37–39 (1973).
5. Tissot, B.P. & Welte, D.H. *Petroleum Formation and Occurrence* 2nd edn (Springer, Berlin 1984).
6. Boreham, C.J. & Powell, T.G. *Org. Geochem.* **11**, 434–449 (1987).
7. Goossens, H., de Leeuw, J.W., Schenck, P.A. & Brassell, S.C. *Nature* **312**, 440–442 (1984).
8. Didyk, B.M. *et al.* *Nature* **272**, 216–222 (1978).
9. Robson, J.N. & Rowland, S.J. *Nature* **324**, 561–563 (1986).
10. Smith, D.J. Thesis, Univ. Bristol (1984).

## Radiocarbon dating

SIR—Radtko proposes electron-spin-resonance analysis as a means of identifying shell and coral samples that are unsuitable for radiocarbon dating<sup>1</sup>. An alternative is to date samples by 'first-order assay'<sup>1</sup>, which provides reliable ages for material less than 10,000 years old and, in so doing, distinguishes between Holocene and earlier samples. Each analysis (excluding labour) costs about £0.50; the requisite liquid scintillation counters are commoner than electron-spin-resonance spectrometers in universities and, of course, radiocarbon laboratories.

C. VITA-FINZI

*University College London,*

*Gower Street, London WC1E 6BT, UK*

1. Radtko, U. *Nature* **333**, 307–303 (1988).

2. Vita-Finzi, C. *Earth planet. Sci. Lett.* **65**, 389–392 (1983).

## Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

# Getting sense out of sequence data

Gunnar von Heijne

*Molecular biologists depend upon computer power for DNA and protein sequence analysis. What should they expect from their software?*

IMAGINE the chief curator of the European Molecular Biology Laboratory's DNA data library arriving in his office one bright morning to find on his desk a note to the effect that the Almighty has deposited, as a personal communication and in computer-readable form, three giga-bases of DNA sequence, comprising the full inventory of the human genome. In His great wisdom, and in an attempt to boost the status of computers in molecular biology, the Almighty has not, however, included any annotations with the sequence data: "Thus", His note informs the chief curator, "I will be most interested in observing your attempts at making sense out of My gift".

Not all of us are chief curators of such databanks, and Divine Grace is normally bestowed upon us in a more piecemeal fashion such as in the form of the occasional successful sequencing gel. Even so, computers are now necessities in molecular biology laboratories, not only for storing and managing restriction analysis and sequencing data, but also for a whole range of further analyses without which little biologically relevant information can be extracted from the DNA and protein sequences. And as the market grows, so does the number of commercially produced software packages. My intention here is to give an overview of what the current molecular biology software can and cannot do, and in what areas we can expect it to do better in the future.

## Functions

A summary of the functions offered by four widely used software packages, chosen at random from the 8–10 commercial products available for IBM XT/AT computers, is presented in Table 1 (but note that frequent updates quickly make lists such as these obsolete). The basic tasks of entering a DNA sequence — either from the keyboard or from a digitizer — taking its complement, and translating one or more reading frames into amino-acid sequence are similar in all four packages. One also finds a corresponding set of tools for entering restriction analysis data. In addition, more or less sophisticated editors often make it possible to build new sequences by 'cloning' on the computer, and there are sometimes programs for automatic construction of restriction maps from digestion data.

Broadly, DNA and protein sequence

analysis can be divided into four main areas: general statistical analyses (base or amino-acid frequencies); signal searches; similarity ('homology') searches; and structural predictions. In a typical series of analyses one would start out with a DNA sequence and search it for open reading frames which 'look like' they might be protein-coding regions. Possible protein–polynucleotide interaction sites such as promoters, splice sites, ribosome-binding sites and RNA hairpin loops are mapped. The DNA sequence or the derived protein sequence is compared to other known sequences or to a whole databank to check for similarities that might reveal something about its biological function or pinpoint conserved structural motifs. Finally, the protein sequence is subjected to various structure prediction schemes.

This may sound simple. But the development of methods and algorithms for efficient and reliable sequence analysis has already reached the point where specialized journals are published<sup>1</sup>, books for both *aficionados* and computer-illiterates are written<sup>2,4</sup>, and software packages containing many tenscores of different programs are marketed<sup>5</sup>. Nevertheless, most of the available methods are far from achieving a satisfactory level of predictive accuracy, and (lest we forget) even though it often takes a computer to make sense out of sequence data, computers can just as easily produce a lot of very impressive nonsense.

DNA sequences are replete with signals such as binding sites for polymerases, regulatory proteins and splicing enzymes. More or less conserved consensus sequences are often found to define the bind-

ing sites, but sometimes geometrical peculiarities of the DNA double helix rather than obvious patterns of nucleotides seem to be involved.

## Weight-matrix

Consensus patterns can be searched for in two ways: by looking for matches to a (possibly degenerate) 'signature' sequence, or by scanning with a so-called weight-matrix (see refs 2–4 for details on the methods discussed in this article). The weight-matrix method is the more general one; it involves setting up a matrix where the natural logarithm of the ratio between the observed and expected frequencies of every base in every position in a sample of aligned signals is recorded, and then scanning the target sequence with this matrix (Fig. 1 — see over). Unfortunately, even the best weight-matrix schemes are far from 100 per cent reliable.

Recognition sites based on the local geometry of the DNA chain are less well understood and much harder to define. The geometrical properties that have been suggested as being important in this regard range from small-scale deviations from the standard B-DNA structure (rotation, rolling or sliding of individual base pairs), to cumulative effects of periodically distributed small bends in the helix, and finally to major structural transitions such as from right-handed B- to left-handed Z-DNA. Such geometrical properties are not immediately recognizable in the primary sequence, and often require elaborate calculations to become apparent. The biological significance of these structural parameters is often hard to assess. ▶

**Table 1** Some functions provided by four commercial molecular biology software packages

|                                      | DNASIS/PROSIS | PC-Gene | IBI/Pustell | MicroGenie |
|--------------------------------------|---------------|---------|-------------|------------|
| Release no.                          | 3.0           | 5.16    | 1.8         | 4.0        |
| Basic sequencing functions           | +             | +       | +           | +          |
| Shotgun editor                       | +             | —       | —           | +          |
| Automatic restriction map generation | +             | —       | —           | —          |
| Find optimal oligonucleotide probes  | —             | +       | +           | —          |
| Find coding regions                  | —             | +       | +           | —          |
| Protein 2D structure prediction      | +             | +       | —           | +          |
| Hydrophobicity analysis              | +             | +       | +           | +          |
| Hydrophobic moment analysis          | —             | +       | —           | —          |
| Global RNA folding                   | —             | +       | —           | —          |
| Dot matrix comparisons               | +             | +       | +           | +          |
| Optimal alignments                   | +             | +       | +           | +          |
| Fast databank search                 | +             | +       | +           | +          |
| CD-ROM supported                     | +             | —       | —           | —          |
| Approximate price (\$)               | 3,000         | 3,000   | 1,200       | 3,000      |
| Vendor                               | LKB           | GenoFit | IBI         | Beckman    |



ACGAATTAGGACTTCCGAGGTATCG

|   |      |       |      |       |
|---|------|-------|------|-------|
| A | (10) | -37   | 2    | (-40) |
| C | -3   | (-12) | -7   | -42   |
| G | -10  | -15   | (-8) | -35   |
| T | -6   | 17    | 10   | 18    |

Score = 10-12-8-40 = -50

**Fig. 1** An imaginary weight-matrix for a signal with the consensus sequence ATTT. The calculation of the match between the first four residues of a target sequence and the weight-matrix is shown (the A in position 1 contributes a weight of 10, C in position 2 contributes -12 and so on). Note that the weights as defined in the text have been multiplied by 10.

Given the poor reliability of present schemes for finding local signals in DNA and RNA sequences, a valuable alternative is provided by methods that analyse statistical properties of larger chunks of sequence. Most importantly, protein-coding regions can be located using criteria related to the fact that the 20 amino acids are not all equally abundant in real proteins, and that certain codons are consistently preferred over others in any given organism. Real coding regions thus tend to show a 'codon bias' that is not found in non-coding sequences, and many algorithms exploiting this have been designed.

RNA molecules add a final complication: tertiary structures where non-contiguous parts of the chain come together to form complex stem-and-loop structures. Although there are many algorithms that can provide everything from a straight listing of all possible hairpin loops to the 'best', optimally folded structure of the whole molecule, these are at best mathematical abstractions that need a lot of extra input in the form of known base-paired or single-stranded regions to provide a trustworthy structure. Inspection of a simple dot-matrix plot where a sequence is compared to itself is often an equally efficient way of finding the most likely structures.

## Protein sequences

For protein sequences, current sequence analysis packages are generally provided with facilities for secondary structure prediction and hydrophobicity analysis. 'Signature' or 'fingerprint' searches intended to uncover potential structural elements such as nucleotide binding sites, signal peptide cleavage sites, glycosylation sites, phosphorylation sites, and so on, are sometimes included.

Schemes for secondary structure prediction have a long history. The well-known Chou-Fasman and Garnier-Robson methods, although more than a decade old, are still very much in use. Based on the frequencies with which the 20 amino acids are found in  $\alpha$ -helices,  $\beta$ -

sheets, turns and coil regions in proteins of known 3D structure, both methods calculate the average local helix, sheet, turn and coil propensities for all positions in the protein chain. These as well as other more recent secondary-structure prediction schemes are much less reliable than most users think: on average, only some 45 per cent of all residues are allocated to the correct structural class in four-state prediction<sup>6</sup>. This poor performance is not likely to be much improved unless tertiary, non-local interactions are taken into account, because only about 20 per cent of all penta- and hexapeptides of identical sequence in proteins of known 3D structure have the same conformation<sup>7</sup>.

Hydrophobicity plots (local average hydrophobicity against position) are used to locate stretches of apolar residues likely to be buried in the interior of a protein or in lipid bilayers, or, conversely, to find polar or charged regions that can be expected to be exposed on the surface. Amphiphilic structures (for example  $\alpha$ -helices with one polar and one apolar face or  $\beta$ -strands with alternating polar and apolar residues) show up in so-called hydrophobic moment plots. Here, one essentially calculates the difference in content of polar and apolar residues between the two sides of a piece of regular structure (over 11 helical residues, say), and plots the resulting value as a function of position. These methods are fairly reliable indicators of transmembrane regions in integral membrane proteins and of peptides with amphiphilic properties (lytic peptides).

A less intuitive, more purely statistical approach to protein structure is cluster analysis (sometimes called discriminant analysis). The idea is to characterize a sequence by a set of numbers such as amino-acid frequencies, percentage of residues predicted as  $\alpha$ -helix or  $\beta$ -sheet, maximum hydrophobicity, maximum hydrophobic moment and so on. This set is then compared to similar sets calculated for a large sample of proteins of known structure, and the 'most similar' protein(s) found in this way are taken to indicate whether the original sequence is likely to be mostly  $\alpha$ -helix,  $\beta$ -sheet or mixed. This kind of analysis can be applied to any number of sequence classification problems such as coding/non-coding regions, protein superfamily classification and so on. 'Signal' searches, finally, depend on consensus sequences or weight-matrices representing the pattern one wants to locate.

When a databank search led to the first unexpected discovery of a sequence homology between an oncogene and a growth factor<sup>8</sup>, the market for fast scanning programs must have increased by orders of magnitude. As a result, no sequencing project today is complete without an exhaustive databank search.

Depending on the sophistication of the method used to find possible similarities, the search can be anything from reasonably fast to impossibly slow. In a fast scan there is a small but finite probability that a fat fish will slip through the net; but, if one cannot restrict the set of sequences to be searched, these are the only programs that will do the job. Once a good candidate is found, more rigorous methods that construct 'optimal' alignments can be applied.

## Optimal alignment

The optimal alignment programs match two sequences by systematically inserting gaps in order to increase the number of matching residues. Sprinkling gaps too liberally will obviously lead to a biologically uninteresting alignment; thus, one associates a cost or gap-weight with each gap. There is a class of algorithms that are guaranteed to produce the alignment with least overall cost (number of matches minus total gap costs). However, the alignment produced can depend critically on the gap-weight value. Programs such as these cannot take information from tertiary structure (if available) into account, and thus cannot preferentially place gaps in surface-exposed regions.

Recent attempts to overcome this problem start from a set of homologous sequences that have been aligned on the basis of one or more known 3D structures<sup>9</sup>. Gaps in relatively conserved 'framework' regions and more variable 'loops' are given different weights, thus forcing the alignment of a new sequence to respect the 3D structure of the original set. As the number of known 3D structures grows, collections of aligned sets ('profiles') representing different protein families or domains can be searched for similarities with new sequences of unknown structure.

Finally, in a world of highly sophisticated automatic alignment procedures, one should not forget the simple dot-matrix comparison. Although 'subjective', important similarities can often be easily identified by eye. When a limited number of sequences are to be compared, dot matrices should always be included in the analysis.

Beyond comparing prices and available functions, there are a couple of other characteristics of good software. 'User friendliness' is a catchword that translates into the two basic features 'ease of use' and 'good documentation'. A molecular biology software package should be accessible even to people who have little or no previous experience with computers; thus, it should be menu-driven and not involve having to issue operating system-level commands (all four packages listed in Table 1 work like this).

Documentation is the second important feature. Manuals should not only teach

the user which particular buttons to push, they should also provide reasonably complete introductions to the various functions offered, with particular emphasis on reliability and possible pitfalls. Original references should be given. It is also useful to have at least some of the documentation available as on-screen help messages.

What can we expect to happen in the next few years? For laboratories that intend to do a lot of databank searching, CD-ROM storage (compact disk) will be the only alternative to using mainframe computers. Already, the GenBank databank requires some 50 Mb of storage space, and currently available hard disks will be outgrown in only a couple of years. Most major software packages either now offer or are about to introduce CD-databank search facilities.

Slow but steady progress in the methods of linear sequence analysis (pattern recognition) can be expected. Control signals of various kinds will be identified with better reliability, and we should also see some improvements in the algorithms used for discriminating between coding and non-coding regions. But in the area of protein structure prediction there seems to be little hope of an immediate breakthrough, at least of a kind that would lead to simple prediction schemes suitable for personal computers. Finding homologies to proteins of known structure seems to be the only way to make even rough 3D-structure predictions at present.

Taking a broad view, computer-assisted sequence analysis is being developed into a useful tool for everyone who deals with DNA or protein sequences. Yet despite frequent references to 'artificial intelligence' and 'expert systems' in advertisements for software, these programs cannot yet be used without heavy reliance on the 'natural intelligence' of the biologist. So, if the Almighty really wants to give molecular biology a push, He should not forget to include a well-documented package of Heavenly Software with His DNA databank gift.

My thanks to Michael Baron for useful comments. □

1. Beynon, R., Modelevsky, J., Roberts, R. & Söll, D. *Computer Appl. Biosci.* **4**, 1 (1988).
2. Bishop, M.J. & Rawlings, C.J. (eds) *Nucleic Acid and Protein Sequence Analysis: A Practical Approach* (IRL Press, Oxford, 1987).
3. von Heijne, G. *Sequence Analysis in Molecular Biology: Treasure Trove or Trivial Pursuit?* (Academic Press, New York, 1987).
4. Doolittle, R.F. *Of URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, California, 1987).
5. Rawlings, C.J. *Software Directory for Molecular Biology* (Macmillan, London, 1986).
6. Nishikawa, K. *Biochim. biophys. Acta* **748**, 285-299 (1983).
7. Argos, P. *J. molec. Biol.* **197**, 331-348 (1987).
8. Doolittle, R.F. *et al. Science* **221**, 275-277 (1983).
9. Gribskov, M., McLachlan, A.D. & Eisenberg, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4355-4358 (1987).

Gunnar von Heijne is in the Department of Molecular Biology, Center for Biotechnology, Karolinska Institutet, Huddinge Hospital - K87, S-141 86 Huddinge, Sweden.

## What's the benefit of compact disk?

Gwyneth Tseng, Goff Sargent and Jack Meadows

**Kirk-Othmer Encyclopedia of Chemical Technology**, 3rd edn. CD-ROM version. Wiley. \$895, £580.

ONE OF the main developments in information technology during the 1980s has been the advent of methods by which large amounts of data can be stored in a very limited space. These methods often involve optical techniques for writing and reading data. CD-ROM (compact disk — read-only-memory) is currently the market leader, its predominance being primarily a result of its swift acceptance as a medium for digital audio recording. Although such disks (120 mm across) are not necessarily the ideal size for text storage, mass production has rapidly reduced both their cost and that of players.

As the 'ROM' suffix indicates, the information stored on compact disks is not meant to be added to or changed; it is simply there to be read, in much the same way as that in a printed book. Indeed, many of the CD-ROM texts now appearing — like the disk reviewed here — are based on existing printed material. The drawback is the cost of transferring the information from one medium to another. At present, sales of text on CD-ROM are limited, and the cost to the producer of licensing the software needed to use the CD-ROM can be high. So only high-value material, mainly reference information, is worth producing on disk.

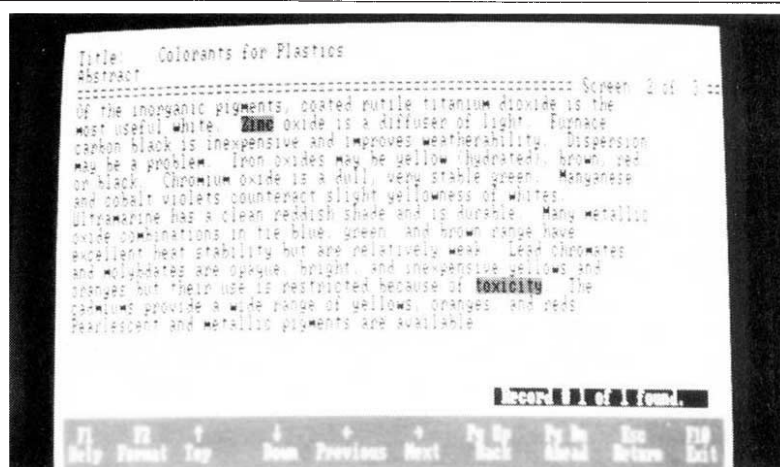
### Hardware

In order to read text from CD-ROM, the player must be linked to a microcomputer which can handle readers' instructions. We use an IBM-PC XT with 640K RAM, a 20 Mb hard disk and a single floppy-disk drive, largely because this meets the advertised requirements of many CD-ROMs. Our original choice of player was the Hitachi CDR3500, a half-height

drive which is designed to slip neatly into the space left below the half-height floppy-disk drive. In the event, the unit supplied did not fit, and we had to swap it for an Hitachi CDR1503S. This is a stand-alone unit, looking much like a domestic compact-disk player.

The Hitachi player can be linked to the PC in 5-10 minutes. An accompanying floppy disk contains software which includes the Microsoft extension, allowing access to the contents of the disks. One snag is that there is, as yet, no general agreement on standards for such things as layout of data on the disk. The 'High Sierra' format (so called from an agreement worked out in the United States) is becoming a *de facto* standard, and this is what is assumed in the Hitachi software. The software is generally easy to use (though placing the floppy in drive 'A:' and typing 'A:SETUP', as suggested, failed to work for us; selecting drive 'A:' and typing 'SETUP' did the job).

The Kirk-Othmer disk proved not to follow the 'High Sierra' format, and attempts to list its files resulted in an error message. The instructions provided were not too helpful in setting up; placing the floppy in drive 'A:' and typing 'INSTALL', as instructed, achieved nothing. Inspection of the disk failed to reveal any such file. There was, however, a file called 'MAKEHD.BAT', and installation proceeded successfully by typing 'MAKEHD HITACHI'. 'MAKEHD' also copied the retrieval software to a specified directory on the hard disk. Rebooting required the CD-ROM to be in place in the drive. Once this was done, and the appropriate directory selected, typing 'FINDIT' led directly to the software,



Kirk-Othmer on-screen — a text page with search terms highlighted.



and so to interrogation of the CD-ROM text.

The advantages of the FINDIT retrieval software that goes with this disk include ease of application and quite good use of colour and windows. The instructions on-screen are straightforward and their execution is fairly quick given the amount of text to be handled. During the search for information, a running count is kept of all the records that match the search requirements. This allows more search terms to be added, so increasing or decreasing at will the number of records retrieved.

## Navigation

The most immediately obvious disadvantage is the complexity of navigating through the text, and especially the tables, on-screen. A number of different keys has to be used, and selecting the wrong one can lead to pages succeeding each other in apparently non-sequential order. A more serious drawback, compared with hard copy, is the lack of diagrams. Although line drawings can be accommodated on CD-ROM, the extra complication has not been thought worthwhile here (nor in the online version). Finally, both the help screens and the documentation relate to the software, not to the database itself, which is where assistance may often be needed.

The appropriate comparison when reviewing text on CD-ROM is not so much hard copy, as online retrieval. Hosts such as BRS, Data-Star and Dialog hold *Kirk-Othmer Online*, along with many other databases. In terms of relative ease of searching, CD-ROM certainly wins. Online searching entails use of a command language, which must be learnt, and use of a communications network with its attendant problems. Then there is the question of cost. Current database costs for *Kirk-Othmer Online* are shown in Table 1. The additional telecommunications charges depend on the place from which the call is made, but CD-ROM must be cheaper for any reasonably high-volume usage of the data.

The disadvantages of online are most apparent when more complex forms of retrieval are wanted. For example, online command languages make it possible to build up phrases such as SOIL SURFACE, or SURFACE OF THE SOIL, which can then be used for retrieval. The software reviewed here would allow retrieval of SOIL AND SURFACE, but the two terms might occur separately in the text, rather than linked in one phrase. These more sophisticated approaches may interest professional information-handlers more than scientists.

The reaction of a chemist was that the CD-ROM version is appreciably less easy to read than the printed version, even though this comes in 25 volumes plus

**Table 1** Costs of using *Kirk-Othmer Online*

|                | Country     | Registration fee | Connect charge per hour | Online print charge per record | Offline print charge |
|----------------|-------------|------------------|-------------------------|--------------------------------|----------------------|
| BRS/Search     | USA         | \$75 per year    | \$80                    | \$0-\$0.37                     | \$0.55 per page      |
| BRS-After Dark | USA         | \$75             | \$22                    | \$0.03                         | n/a                  |
| Data-Star      | Switzerland | None             | \$55                    | None                           | None                 |
| Dialog         | USA         | \$25 per year    | \$85                    | \$0.25                         | £0.55 per record     |

index. The main problem was again charting a course through the text, which was exacerbated by an inadequate indication of where the reader is and by the difficulty of jumping rapidly between non-consecutive pages. All tabular material appears after the accompanying text, which makes continuous reading difficult. Not only the diagrams but the multi-dimensional formulae are omitted, while chemical formulae are represented in a rather unhelpful way by structural fragments separated by spaces (for example, H<sub>2</sub>SO<sub>4</sub>, or H-2 SO-4). Nor is this mode of representation consistently maintained. Scientific notation is not incorporated, so that equations are difficult to read. One useful feature compared with the hard copy is the inclusion of an abstract, also present in the online version.

A drawback of many CD-ROMs currently on the market is that they simply combine data from the existing hard copy and software from the existing online versions. This seems to be essentially what has been done for the *Kirk-Othmer* disk. A better approach might be to rethink needs in terms of the specific capabilities

of CD-ROMs. Meanwhile, such disks are probably best handled by information professionals and subject specialists in tandem, rather than by the latter acting alone.

The CD-ROM market is just beginning to take off, with over 300 titles currently available for purchase. The general belief is that choice of disks will grow quite rapidly, and that compact disks will become the lead product in mass-storage publishing. As our experience shows, there remains the problem of standards, especially those relating to format. If agreement cannot be reached soon, the same competition that occurred with video players may reappear with compact disks. To complicate prediction, new kinds of disk, such as CD-I (compact disk — interactive), are appearing which require further changes in formatting standards. Cautious publishers will, no doubt, continue to hedge their bets for another year or two. □

Gwyneth Tseng, Goff Sargent and Jack Meadows are in the Department of Library and Information Studies, Loughborough University, Leicestershire LE11 3TU, UK.

## Model maker

A.J. Rowe

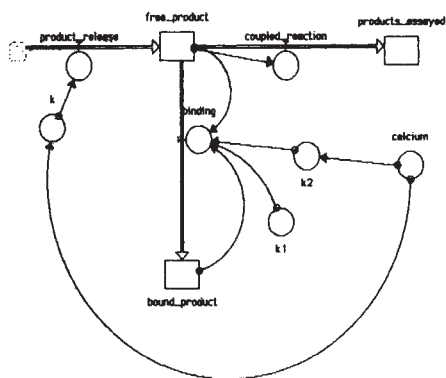
**STELLA 2.0. High Performance Systems, 13 Dartmouth College Highway, Lyme, New Hampshire 03768/Logotron, Dale's Brewery, Gwydir Street, Cambridge CB1 2LJ. \$150, £199 + VAT. STELLA 2.0 for the Macintosh II, \$225, £299 + VAT.**

COULD it be that the contemporary fragmentation of knowledge, so much lamented, is not entirely one-way traffic? *STELLA* for academe (there is also a version for business) tempts one to suppose so. Devised by Barry Richmond, a professor of engineering, it is a modelling package for the Apple Macintosh and can be applied in economics, resource management, behavioural studies, a whole range of experimental sciences (including population biology, reaction kinetics, geology) — and much else besides. The current version of this remarkable program, even more user-friendly than the original, will surely cause many to give thought to the notion that a common heuristic

underpins problem-solving across many disciplines.

Jay Forrester's 'system dynamics' forms the conceptual framework for *STELLA*. Describing systems which are subject to time-dependent change in terms of a limited number of phenomenological parameters with appropriate linked coefficients is, of course, hardly new. The thermodynamics of irreversible processes comes to mind. What computer scientists (or just plain enthusiasts) are now doing, however, is to make available, across the range of disciplines, approaches that were previously restricted in application not so much by academic need as by the user's ability to cope with and implement the mathematical algorithms involved.

Although software packages of statistical routines have been popular for many years, their use has been passive. Users collect their data, ideally with some regard for the format required and the nature of the questions to be answered, and then analyse them through the package. *STELLA* is different. Users construct models themselves, employing the range of simple but powerful tools provided. Creative input is necessary, and a great deal more thought is called for, but the



**STELLA** diagram modelling an enzyme-catalysed system. The product, after release from the active site (rate constant  $k$ ), can either be adsorbed onto excess low-affinity sites on the enzyme surface (with rate constants  $k_1$  and  $k_2$  for the on and off reactions respectively) or enters a coupled assay system (non rate-limiting). Both  $k$  and  $k_2$  are modulated by the level of calcium ions. The model enables the effect on the coupled assay of addition/removal of calcium at steady state to be explored.

outcome can be immensely valuable and informative.

**STELLA** uses the familiar windows/icon/mouse/pull-down menus interface. Structures are generated on the Macintosh screen as a (flow) diagram, using just four basic icons which denote stocks, flows between stocks, convertors (factors which influence flows) and connectors (lines denoting a logical connection between a stock or convertor and a flow). In the case of a chemical reaction, for example, these could correspond to reactants/intermediates, reaction pathways, catalysts and linkages of catalyst to the step catalysed, respectively. Input screens then enable the user to enter the parameters to be assumed. It is even possible to do back-of-an-envelope modelling, by sketching out assumed relationships on a graph pad. The program numerically solves the appropriate differential equations, and displays results as graphical or tabulated data as the user requires.

A newcomer to **STELLA** (if not to modelling), I found it easy to learn how to use the package. One could quibble over minor things. The immense user's guide of 392 pages is wordy and uneven — though the vital bits are clear. Mathematicians will find the choice of numerical methods limited, and I was disappointed that the graphics window could not display the main parameters assumed. No matter. This is a package which may well do for modelling what BASIC (also from Dartmouth, New Hampshire) did for programming — make it accessible to the many and lead at least some onto more complex work. **STELLA** will surely be widely used. □

A. J. Rowe is in the Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK.

## Information from data

Edwin Metcalfe

**CLEOPATRA: A Chemometrics Library.** Elsevier Scientific Software. Basis version Dfl. 2,220, \$1,080. Extension 1 Dfl. 1,010, \$495.

CHEMOMETRICS can be broadly defined as the application of information technology to the analysis of chemical data. The essential goal is to process and convert data into information as efficiently and effectively as possible. Good experimental design provides the maximum amount of information from the minimum amount of experimentation. Other chemometric methods are aimed at obtaining the maximum amount of information from large and often complex data sets.

**CLEOPATRA** is a modular chemometrics teaching package which, although it can't be said to be comprehensive, covers several important areas of the subject. A lot of thought has gone into the presentation of the material, which has a modular structure, allowing individual units to be selected by menu. The package is available for both the IBM-PC and HP microcomputers. I have used the IBM version on four PC clones without any problems, provided (even though the program does not use colour) that the microcomputer has a colour graphics adapter card. A particularly useful feature is that the program disks are supplied in two sets, one of interpreted BASIC and one of compiled BASIC. With the interpreter version, the user can incorporate his own BASIC teaching programs as extra modules or modify the existing modules. Some of the modules involve a great deal of calculation for which the faster compiled version is preferable.

Nine modules come in the 'basis' version, including Fourier filtering, curve-fitting and autocorrelation, with six modules covering demonstrations and calculations in sampling, process monitoring and process reconstruction. An extension is available for simplex optimization, Kalman filtering and sequential analysis. The package is unusually well documented — the handbook clearly describes how to use each module and outlines the theory, though newcomers to the subject will need to refer to a textbook or the references given on the header page of each module.

I particularly liked the graphic presentation of the data. In the Fourier-filtering module, for example, as a spectrum is acquired the signal varies slowly (the peaks are broad), while the noise varies rapidly (narrow peaks), and the signal and noise are superimposed in the spectrum.

But on applying the fast Fourier transform algorithm, the data are converted to a frequency or power spectrum where the signal and noise are separated. A filter can cut out the high-frequency noise before carrying out an inverse transform to regenerate the smoothed spectrum. The screen is split into four, showing the original noisy spectrum, the power spectrum, the shape of the filter used and the reconstructed spectrum.

There is a great deal of flexibility in the data used; up to three peaks with varying levels of noise and sinusoidal distortion, and several different types of filter, can be investigated. Similarly, in the curve-fitting module, up to three peaks of different height, position, width and shape (gaussian, lorentzian or mixed) can be studied with various levels of noise superimposed. The Kalman filter, a recursive method which can be applied to multi-component spectroscopic analyses, can be used with the overlapping spectra of four components of variable concentrations and in the presence of baseline drift and variable amounts of noise.

Such flexibility is one of the most important features of any good teaching package. It makes exercises open-ended and more challenging by enabling students to ask such questions as how well can the Kalman filter or curve-fitting routines cope with a high level of noise, or what happens if two peaks are fitted to data which also contain a third underlying peak.

The simplex optimization module provides an excellent visual demonstration of how an optimum response can be obtained for two variables, essentially by climbing to the top of the hill and representing the response by the steepest path. In practice most simplex optimizations will use far more than two variables, but then visualization in more than three dimensions is not possible.

In my experience students enjoy using **CLEOPATRA**, provided care is taken to give them both guidance and the freedom to experiment with open-ended exercises. The main frustration arises from the self-contained nature of the package; with the exception of the calculation modules on grab sampling, process monitoring and process reconstruction, the data are internally generated and the modules cannot be used with the students' own experimental results (with a few modules it is possible to modify the programs, for example the Fourier filtering and curve-fitting units, to allow data files to be read from disk). This drawback does not detract greatly from the usefulness of the package as a teaching aid. In my view, **CLEOPATRA** is an essential component of any chemometrics software library. □

Edwin Metcalfe is in the School of Chemistry, Thames Polytechnic, Wellington Street, Woolwich, London SE18 6PF, UK.



# The molecule in print

Gary Ashley

**ChemText 1.2.** *Molecular Design*, 2132 Farallon Drive, San Leandro, California 94577/Armstrong Mall, Unit B10, Southwood Summit Centre, Farnborough, Hampshire GU14 0NR. \$1,500, £1,500 (industry); \$500, £750 (academic).

*ChemText* is an integrated chemical structure-drawing and word-processing package that allows the user to produce chemical documents on an IBM-PC computer or compatible. Unlike most other chemical software packages, text and graphics are controlled within one program and can be displayed simultaneously on the screen, allowing great ease and flexibility in document design.

While its word-processing capabilities compare favourably with 'standard' programs, *ChemText* is unusual among IBM-PC word processors in that it has been designed with the non-technical user in mind. It is almost exclusively mouse-driven with pull-down menus, so that there are no complex control codes to memorize or search for on a keyboard template. The program has a hierarchical

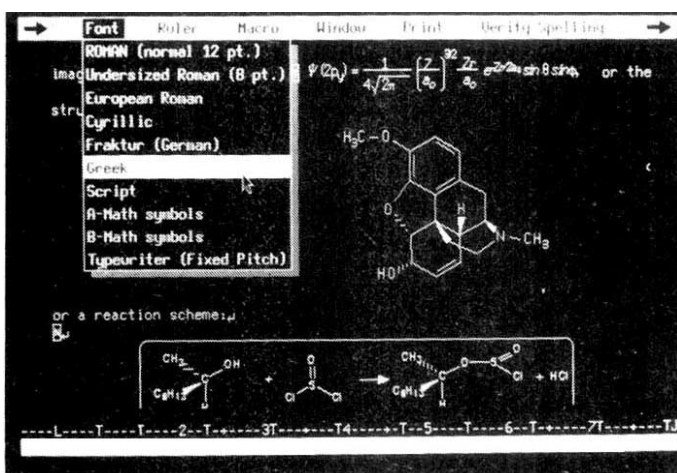
structure, separating the three basic tasks of molecule drawing, sketch editing and word processing into separate modules.

Individual molecules are drawn and manipulated at the innermost level of the program. *ChemText* provides several commonly used partial structures, or templates, to aid in this process, and allows for storage of user-defined templates. Stereochemical bonds are readily indicated, and bonds may be highlighted for perspective views. The program automatically checks valence, and the user can choose to display no hydrogen atoms, only those hydrogens attached to heteroatoms, or those attached to terminal carbons as well, with the program calculating the correct number of hydrogens at each position. There are also helpful features to indicate charge and isotopic composition, and a 'clean-up' facility which evens out structures drawn freehand.

At the next level up, molecules are assembled into 'sketches', complete with arrows, arcs, explanatory text, boxes and other notations. Molecules may only be moved and re-sized at this level. Alignment in complex diagrams is aided by movable, ruled crosshairs, a feature often lacking in structure-editing programs. A notable advantage of this division of labour is that once a molecule is drawn in the lowest level, it is treated as an object to be manipulated in the sketch editor. The user can draw a molecule without regard to size or placement in a diagram, and then fit various structures together at will to form a complex diagram.

The highest level of the program is the word processor, where the sketches can be combined with text to produce a completed document. *ChemText* allows great flexibility in page layout, enabling drawings to be placed in line with the text or as block diagrams. Page spacing is calculated automatically, and diagrams too large to fit the remaining space on a partially filled page are moved to the next page; the user may select either to have the program fill in the resulting space with text or to leave it blank.

The word-processing part of the package has most of the features found in standard programs, such as macro capability, windowing, block commands, headers and footers, text string location, and spelling checking using either American, British, German or French dictionaries. Because the entire screen is displayed in graphics, *ChemText* shows



On the menu — *ChemText*'s choice of fonts pulled down onto work in progress on the screen.

subscripts, superscripts, boldface, italics and special font characters as they will be printed, without having to replace the video character PROM chip in the computer. A variety of fonts are available, including Roman, undersized Roman, European Roman (Å, ä, for example), Cyrillic, Fraktur, Greek, script and various mathematical symbols.

The true test of any software package of this type is its output quality. *ChemText* is compatible with a wide range of printers, including those of the dot-matrix and laser type. For superb 300 dots per inch output, a PostScript printer such as the Apple LaserWriter is needed, although quite good 150 dots per inch is obtained with non-PostScript printers. *ChemText* supports 300 dots per inch of full-page graphics on a Hewlett-Packard LaserJet Series II having 2 Mb of memory; otherwise, only 32 square inches of 300 dots per inch graphics per page can be supported.

Although the program will theoretically run on any PC family member, various practicalities limit its effective use to the PC-AT. The power of full-screen graphics has its price — the program takes up nearly 2 Mb of disk space and cannot be run from floppy disks. The extra speed of the 80286 is really needed. Even on a 10 MHz, zero-wait AT system, a pause long enough to be annoying accompanies image-drawing during scrolling in the text editor, and one might well die of old age during this process on a PC-XT.

*ChemText* is a top-of-the-line word processor for chemists. It is easy to learn and use, making it well worth its considerable expense. The program combines the best elements of the Macintosh user interface with the attributes of the IBM-PC AT. In terms of power and ease of use, there are really no software packages available which can compare with this one. □

Gary Ashley is in the Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA.

## ADVERTISEMENT

# The ANS Journals







Your number one source for current information about

## NUCLEAR ENERGY

**NUCLEAR NEWS**

**NUCLEAR SCIENCE & ENGINEERING**

**NUCLEAR TECHNOLOGY TRANSACTIONS**

**REMOTE SYSTEMS TECHNOLOGY PROCEEDINGS**

**FUSION TECHNOLOGY**

For free catalog contact:

**AMERICAN NUCLEAR SOCIETY**  
 555 N. Kensington Avenue  
 La Grange Park, Illinois 60525 USA  
 312-352-6611 Telex: 4972673

# The need for standards in image processing

Kendall Preston, Jr

*Image processing is crucial to many disciplines. But lack of standardization of the software involved means that the wheel is continually being reinvented.*

IMAGES are processed in everything from simulation graphics to computer vision and human perception. When computerized, the range of applications is vast, and it is no wonder that there is no universal software package for the job. There is software for processing images produced by spacecraft; for Earth resources images; for radiological images; for images generated in microscopy; for factory production-line robot vision; and for graphics. Unfortunately, in this diversity there is also massive redundancy.

The radiologist who sits, hand on tracker ball, in front of the CT (computer tomography) screen is involved in processing images. The tracker ball is used to change the 'level' and 'window' settings in order to modify the average value and the range of values of the CT numbers which form the radiological image on the screen. Because a computer is used for this purpose, this is digital image processing. The electron beam of the television tube is modulated via a digital-to-analogue converter by an amount proportional to the CT numbers stored in computer-like 'refresh' memory.

## Multispectral data

By the same token, image processing is performed by the computer which combines the multispectral data gathered by an Earth-observation weather satellite to generate the colour display seen in the evening weather news report. Special 'graphics' embellishments may be added to this display to represent more clearly the flow of the jet stream or the position of warm and cold fronts. But perhaps the most exquisite form of digital image processing is that conducted by the mini-supercomputers installed in the robot microscopes used in haematology for the white blood cell differential (WBCD) count. These extraordinary machines use special television cameras operating at 100 frames per second to review images of one hundred thousand human red cells while searching for the 100 white cells whose images are digitized and computer analysed to measure the features needed to produce an automatic WBCD.

Thus processing is used not only to generate displays of image data to be analysed by the human eye but also to permit some of these analyses to be executed automatically. In the latter cases the digital image processor replaces the human eye and brain.

One of the barriers to standardization is

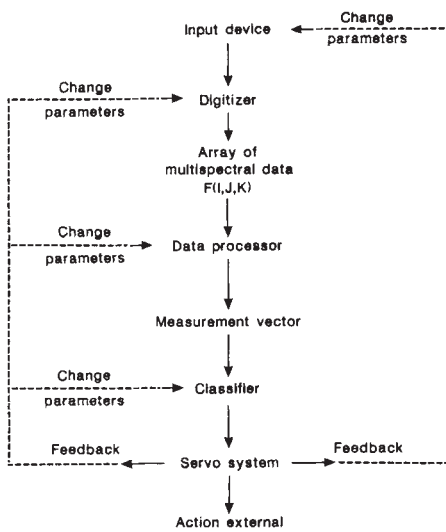


Fig. 1 Flow diagram for image processing and recognition.

the complexity of the image-processing environment. Figure 1 is a schematic representation of a complete image-processing system. At the top is the input device which, via a digitizer, stores image data. There are innumerable types of imaging device, making standardization of data format all but impossible. Once digitized, the multispectral information may then be transformed and processed to yield a display for human observation. Alternatively, a measurement vector may be generated and then shipped to a classifier for recognition purposes. So image processing for display assumes human

involvement while image processing for recognition is the automation of human vision. Feedback may occur within the system to change the parameters of the input device, the digitizer or the processor. When this feedback is adaptive and 'learns by experience' then the system is said to have artificial intelligence.

Clearly, standardization of a total image-processing software package is a formidable task, which no doubt explains why it has yet to be achieved. At Carnegie-Mellon we have reviewed over 100 packages and have tabulated 3,000 subroutines from those for which documentation is available<sup>1</sup>. This has led to a useful method of subroutine categorization (see Table 1).

Here it is worthwhile digressing for a moment to look at subcategory 2.3, computer graphics. Because the operations entailed in plotting, drawing and shading are well defined, computer-graphics standards have evolved. Early on, in the 1960s, CalComp (California Computer Products) wrote a set of Fortran-callable subroutines which became the *de facto* standard for drawing and plotting. The GCS (Graphics Compatibility Standard) was then formulated at the United States Military Academy in the early 1970s and was promulgated by the United States Army Corps of Engineers. The Special Interest Group on Graphics of the Association For Computing Machinery (ACM/SIGGRAPH) was formed and moves towards standardization began in April 1974 at a meeting entitled "Device-Inde-

Table 1 Categories and subcategories of image-processing subroutines

- |                              |                                                    |
|------------------------------|----------------------------------------------------|
| 1. Utilities                 | 4.3 Mosaicing/registration                         |
| 1.1 Identifiers              | 4.4 Map projection                                 |
| 1.2 Executives               | 4.5 Gridding/masking                               |
| 1.3 Formatters               | 5. Image transforms                                |
| 1.4 I/O Commands             | 5.1 Noise removal                                  |
| 1.5 Test patterns generators | 5.2 Fourier analysis and other spectral transforms |
| 1.6 Help files               | 5.3 Power spectrum                                 |
| 2. Image display             | 5.4 Filtering                                      |
| 2.1 Cathode ray tube         | 5.5 Cellular logic                                 |
| 2.2 Hard copy                | 6. Image measurement                               |
| 2.3 Interactive graphics     | 6.1 Histogramming                                  |
| 3. Arithmetic operators      | 6.2 Statistical                                    |
| 3.1 Point                    | 6.3 Principal components                           |
| 3.2 Line (vector)            | 7. Decision theoretic                              |
| 3.3 Matrix                   | 7.1 Feature select (training)                      |
| 3.4 Complex number           | 7.2 Classify (unsupervised)                        |
| 3.5 Boolean                  | 7.3 Classify (supervised)                          |
| 4. Geometric manipulation    | 7.4 Evaluate results                               |
| 4.1 Scaling/rotation         |                                                    |
| 4.2 Rectification            |                                                    |



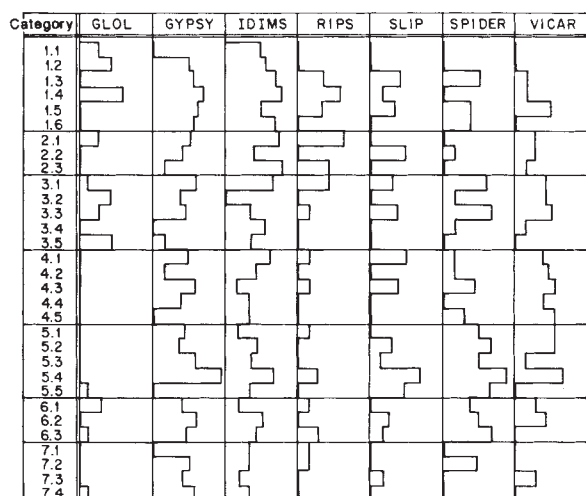


Fig. 2 Plot (unity-biased logarithmic scale) of the signatures of seven image-processing packages showing the number of subroutines in each category.

pendent Computer Graphics" at the United States National Bureau of Standards. This work was further supported by IFIP (International Federation for Information Processing) and ISO (International Standards Organizations).

In image processing the story is not such a happy one. In May 1977 a workshop was held on standards for image pattern recognition (that is, for image processing)<sup>2</sup>. Although the result was the establishment of the NASA CCT (Computer Compatible Tape) format for multispectral Earth observation sensor data and the NATO RSG-4 format for the exchange of digitized image data (as developed by the NATO Sub-Group on Image Processing), it did not lead to any coherent steps towards general standardization in the United States.

But what specific software is available? Japan has been the first country to succeed, at least nationally, in developing a standard transportable package. Starting at Nagoya University with the work on SLIP (Standard Library for Image Processing) and the subsequent evolution of this package into SPIDER (Subroutine Package for Image Data Enhancement and Recognition), a set of approximately 400 Fortran-callable subroutines were generated as the workhorse package. SPIDER is used by most university and industrial laboratories in Japan but has not been adopted worldwide.

In the United States, KANDIDATS (Kansas Digital Data System) was produced at the University of Kansas in the early 1970s, after which came GYPSY (General Image Processing System) from the same hand but at Virginia Polytechnic Institute. GYPSY is now used in many parts of the United States; originally written as a transportable package, it has recently been recast for the Sun Microsystems work-station to run under UNIX, and will soon be available on the Cray

supercomputer.

Other packages for image processing used in the United States are specific to particular areas<sup>3</sup>. In meteorology, for example, there is the massive McIDAS, developed by the University of Wisconsin. This package (over two million lines of source code), originally coded for Harris computers, has recently been ported to personal computers running under the OS/2 operating system. Under a National Science Foundation programme, 1,000 users are anticipated nationwide over the next few years.

For the analysis of images from spacecraft there is VICAR (Video Information Communication and Retrieval)

devised at the California Institute of Technology Jet Propulsion Laboratory. A minicomputer version, Mini-VICAR also has been written. In the closely related field of Earth resources image analysis there is IDIMS (Interactive Digital Manipulation Systems) from the Environmental Systems Laboratory (Sunnyvale), ELAS (Earth Resources Laboratory Application Software) from the National Space Technology Laboratories, IDAPS — now incorporated in EADS (Engineering Analysis Data System) — from the Marshall Space Flight Center and RIPS (Remote Image Processing System) from the EROS Data Center of the United States Department of the Interior.

In medicine there is TICAS, developed at the University of Arizona for image analysis in cytology; GLOL (Golay Logic Operating Language), used in the WBCD counting robots developed by Coulter Biomedical Research; and PL/S, developed at the Massachusetts General Hospital for radiology.

Similar, highly selective, lists have been compiled for Europe. Those packages mentioned above are used at Carnegie-Mellon to identify the 'signatures' of image-processing software. These signatures are useful in evaluating the capabilities of a specific package (see Fig. 2). The numbers in the left-hand column of the figure refer to the categories and subcategories given in Table 1. Signatures, plotted on a unity-biased logarithmic scale (to avoid negative numbers) show the number of subroutines in each of the subcategories for a particular language. Although no package has total capabilities, IDIMS shows only a single null (in subcategory 3.2, vector operations). GYPSY has only five nulls, while other packages have many. The correlation between SLIP and SPIDER is clearly evident. In using these signatures the user should first graph the signature which

corresponds to his requirements and then correlate it with the signatures of available packages in order to find the best match.

Unquestionably, there is an oversupply of image-processing software. The forces which drive its development come from researchers in diverse disciplines. Communications between working groups are poor. Although some user groups exist, there is no society devoted to the subject. It is encouraging to find a concerted effort towards standardization in Japan, yet hundreds of man years are annually wasted worldwide in devising one 'new' image-processing package after another. At times, it seems that we are on a treadmill going nowhere — too often I have seen splendid applications routines written using one image-processing package so locked into its parochial system that researchers at other universities, who wanted to use the code, found it untransferrable. This is more than a pity.

## Machine specificity

A further difficulty which compounds standardization problems is the machine specificity of many packages. VICAR, for example, consists of 150,000 lines of source code of which 50 per cent is in assembly language for IBM mainframes. Furthermore, speed is so important that many packages run on specific peripherals. For example, the EADS routines have been written for the Image International Systems Model 75 image processor. Such a package, although of great value to certain users, is useless for those not possessing the required hardware.

What are the prospects for the future? This year, a new ANSI X3H3 technical subcommittee has been formed with the object of producing an image-processing kernel by 1990. The work of this subcommittee should encourage transfer of software from one group to another, help establish benchmarks for testing and comparing packages, and — eventually — catalyse the formulation of true standards. Close coordination with existing ACM/SIGGRAPH efforts in setting standards for raster graphics will be a necessity, but then coordination is the way to win in this particular game.

I thank Todd Greer, Mary Adams and Joanna Damhesel for help in compiling this article, and the UK Medical Research Council, National Research Council of Italy and the US National Science Foundation for support. □

1. Preston, K. Jr in *Computing Structures for Image-Processing* (ed. Duff, M.J.B.) 195–211 (Academic, London, 1983).
2. Evans, J.M. Jr, Kirsch, R. & Nagel, R.N. (eds) *Workshop on standards for image pattern recognition* (NBS Spec. Publ. No. 500–8, May, 1977).
3. Chang, S.-K., Ichikawa, T. & Ligomenides, P. A. *Visual Languages* (Plenum, New York, 1986).

Kendall Preston, Jr is in the Department of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA.

# The art of scientific programming

Colin Upstill

**Numerical Recipes in C.** By William H. Press, Brian P. Flannery, Saul A. Teukolsky and William T. Vetterling. Cambridge University Press: 1988. Pp. 735. £27.50, \$44.50; diskette £15, \$29.95. Example book £15, \$19.95; diskette £15, \$24.95.

IN remarkable contrast to scientific computers, scientific programming has changed little in a quarter of a century. A program written in current standard FORTRAN would be quite recognizable to a programmer from the 1960s. Similarly, Pascal, although sometimes referred to as a third-generation language, is a close relative of the venerable Algol. No scientist would want a first-generation vacuum-tube computer, but the scientific computing community seems unwilling or unable to move on from first-generation programming languages.

Programmer productivity is approximately constant at about ten lines of debugged code per person per day, irrespective of the level of the language. This alone is a powerful argument for higher-level languages. Yet although most scientists make use of computers to some degree or other, they continue to allow a state of affairs in which the common scientific programming languages do not contain any appropriate constructs for matrix manipulation, calculus, vector analysis and many other examples of the sort of mathematics which is regularly needed in scientific computing.

In the early days of electronic computers, programming was possible only in machine code. Although this closeness to the machine sometimes has its advantages, all but the most trivial of tasks are exceedingly tiresome to effect. In those early days, the hardware was the most important factor because of its relative cost, and immense programming time and effort was invested to maximize use of the machine. As the power of computers increased, so did the complexity of programs, and the cost of programming began to equal the cost of the computer. For this reason, and simply for practicable programming, it became essential to use higher-level constructs which are translated into the verbosity of simple instructions the machine understands by a program written for that purpose.

The name FORTRAN is a contraction of FORMula TRANslation, a choice which reflects the original emphasis on producing a program which translated formulae into machine code: the design team's target was for a machine-translated program to execute in no more than twice the time taken by a hand-coded program. The translator (compiler, in modern parlance) was the challenge, not the design of the high-level programming language in which the formulae were to be written.

Thus it was that the earliest dialect of FORTRAN was quite casually conceived, features being included or excluded according to their ease of translation into IBM 704 machine code rather than from programming considerations.

In software engineering, much is made of efficiency. This usually refers to the hardware and software, but human resources should not be neglected. Many pocket calculators are easier to program than their much more powerful brethren,

**"Programmer productivity is approximately constant at about ten lines of debugged code per person per day, irrespective of the level of the language. This alone is a powerful argument for higher-level languages."**

and can be more effective when what matters is simply getting a result in the minimum time. What should a scientist care if the departmental computer can do his calculation in seconds, if it takes many minutes of his time to achieve a working program, when his pocket calculator can be programmed in a couple of minutes? Obviously the costs and benefits are different when the same program is to be run many times, but that does not explain why there has been so little advance in scientific programming languages.

There is of course no simple answer. Politically, it is difficult to get agreement upon the content and syntax of a language which can provide enough well-chosen high-level constructs to satisfy the diverse requirements of scientific computing. A different compiler is required for each type of machine, and the cost of each is considerable: clearly this investment will not be made until the political problems are resolved. The prospects for that are bleak. The long-overdue revision of FORTRAN, known as FORTRAN 8x because it is (was?) expected in the 1980s, should be a step in the right direction, but the project has been much delayed.

With the advent of affordable parallel processing based on multiple microprocessors, a new generation of much more powerful machines is imminent. Such machines demand new high-level languages, with and without parallel constructs. Those with parallel constructs are necessary when the programmer needs to have control over the way the machine partitions a problem. But he thinks he needs such control far more often than he really does, and in most cases what is

needed is a high-level way of stating the problem and leaving the machine and the system software to get on with the job: too low a level of problem statement maximizes the danger of programmers' preconceptions forcing inefficiency on the machine.

It is the need for access to facilities which should dictate the programming language used, but all too often it is simply fashion, or subliminal pressure from computer 'experts' who revel in their intimacy with the machine. These phenomena have led to the current excess of misplaced enthusiasm for the C programming language. Although its most notable protagonists introduce C as a general-purpose programming language, it was devised originally for systems programming and puts the programmer close to the machine in several respects. For that purpose, C is excellent — indeed, it is the underlying language of the UNIX operating system. Its value for scientific programming is far less certain.

Of the many high-level fourth-generation languages proposed, none looks like gaining wide acceptance and enduring, save possibly in business computing. But there, with spreadsheets and the like, we are generally in the grey area between languages and applications programs, and it is arguable whether one is a programmer or a user. It is another region of this grey area which rescues current scientific programming, and that is the existence of libraries of subprograms which perform particular high-level operations under the control of lists of arguments. These are available from many sources, but there is danger in using them without detailed evaluation of their internal workings: a proper understanding of the mathematics underlying a subprogram is essential if the common folly of applying numerical methods inappropriately is to be avoided. Other problems include the degree of completeness of coverage, the cost of fully warranted and supported libraries, the availability of modifiable source code if something just a bit different is required, and (in the public domain especially) the provenance, trustworthiness and efficiency of library subprograms.

*Numerical Recipes: The Art of Scientific Computing*, which appeared last year, went a long way to overcoming many of these problems. It provides, in one volume, the description and source code of a comprehensive library of scientific subprograms, merged with a numerical analysis textbook sufficient to leave few excuses for misapplying them. (Moreover, for those who require very many of the routines, or for the terminally lazy one-fingered typist, the software is available cheaply in machine-readable form.) Far from being yet another dull programming manual crossed with a pot-boiler of a numerical analysis text, the book is



extremely readable, in a style which owes an acknowledged debt to Acton's *Numerical Methods That Work*, one of the few books on numerical analysis which could be described as hard to put down.

It is an indication of the penetration of C that the authors have seen fit to produce a C language edition of their splendid book, to be sold alongside the existing combined FORTRAN and Pascal edition. The numerical recipes are the same, and most of the text is identical. The textual differences are telling, in that they are almost all additions to describe ways of handling the shortcomings of C for scientific programming, including a ten-page

**"It is the need for access to facilities which should dictate the programming language used . . ."**

discussion of ways of circumventing intrinsic problems with array manipulation, the lack of complex variables and complex arithmetic (for which routines are supplied in an appendix), and the problems of controlling conversion between single and double precision. The publisher claims that this edition "brings to bear on scientific tasks the full power of the operator-rich C language, including dynamic memory allocation, modularization, pointer reference to matrices, structured programming, and much more". Modularization and structured programming are perfectly possible in FORTRAN, Pascal, and many other languages; the remaining attributes of C are things that most scientific programmers would rather not be troubled with.

Throughout the C edition, it is noticeable that, largely as a consequence of uncomfortable array manipulations, the programs are harder to read and to relate to the mathematics than their FORTRAN or Pascal counterparts. Besides the inconvenience of writing scientific programs in C, programs need to be understood if they are to be modified — and most programs need to be modified at some time or other. On these grounds alone, C falls short. Moreover, given that on most machines there is little difference in performance whether we use C, FORTRAN or Pascal, I can see no reason to use a language which is manifestly worse for a particular purpose if a better one is available. It is the use of the language for other than its intended purpose which has brought FORTRAN opprobrium it has not always deserved.

The last thing I desire is that my opinions about scientific programming should be construed as condemnation of this book. If, for whatever reason, you do your scientific programming in C, then buy it — and if you do not, then get the other edition. □

Colin Upstill is with Plessey Research, Roke Manor, Romsey SO51 0ZN, UK.

## Money matters

Sharon Cohen

**Grant Manager 2.0/Personnel Manager 1.4.** Niles and Associates. 2200 Powell, Suite 765, Emeryville, California 94608. \$425 (if purchased separately), \$740 (combined).

APPLYING for grants, budgeting grant funds and accounting for how the money is spent are not pleasurable activities. Computers should accelerate and simplify grant management. Until recently there were a limited number of effective software programs for this purpose.

*Grant Manager* is an accounting program expressly designed for the management of research and training grants, contracts and departmental budgets. It is of value to individual researchers with only a few grants — but from several different sources — and to departmental administrators who must monitor many types of grant. *Grant Manager* can provide the answers to such questions as: Does my National Institutes of Health grant have any funds remaining in the equipment category? How much money do I have in each of my grants and contracts as of today? Where do I usually buy caesium chloride? How much do I normally spend

order form.

*Grant Manager* can perform searches for orders or specific items, and can list how much has been spent on supplies, personnel, radiation protection or other costs over a specified period of time for each of the accounts entered. When chemistry supplies are running low, a simple search allows one to find out when the last batch was ordered, how much it cost, which vendor was used and the catalogue number. This feature allows for easier budgeting, and saves hours of rummaging through old files.

The program maintains current balances on all accounts and simplifies ledger reconciliations. A new balance is created each time the data are amended. The ledger reconciliation module shortens the time required to match ledger items to the program's records. Overdrafts and unspent funds left over from a particular budget period should be a thing of the past.

Tracking vendor histories is vital for those involved in purchasing. The program provides details of gross costs, delivery dates and volume, thus making negotiation of discounts easier.

*Personnel Manager* complements *Grant Manager* and allows personnel costs to be included in the grant-tracking process. By allocating such costs for the entire budget period, researchers and administrators

| GRANT MANAGER<br>ITEM SOURCE |                                          |          |
|------------------------------|------------------------------------------|----------|
| VENDOR                       | ITEM DESCRIPTION                         | ORDERED  |
| AMERICAN SCIENTIFIC PRODUCTS | 25330-50 Corning Centrifuge Tubes 50ml   | 07/04/85 |
| AMERICAN SCIENTIFIC PRODUCTS | 25310-15 15ml Centrifuge Tubes           | 07/04/85 |
| BECKMAN INSTRUMENTS          | #342414 Quick seal tubes for VT50 rotor  | 05/17/85 |
| BECKMAN INSTRUMENTS          | #344090 Ultra Clear Tubes 3/16 x 1 5/8"  | 05/17/85 |
| BECKMAN INSTRUMENTS          | #344057 Ultra Clear Tubes 1/2 x 2"       | 05/17/85 |
| BECKMAN INSTRUMENTS          | 326823 Polyallomer Tubes (3 1/2"x1")     | 05/31/85 |
| BECKMAN INSTRUMENTS          | 337986 Polyallomer Tubes 5/8"x4"         | 06/05/85 |
| GENINT SCIENTIFIC COMPANY    | 505 Micro Centrifuge tubes 1.5ml 500/bag | 06/01/85 |

*Grant Manager* search of previously ordered items. Here the search term was 'tubes'.

on supplies each month? Can I afford to hire a technician? Over the past three years I have been getting answers to these questions for 30 faculty members with 135 grants. The results have surpassed all expectations.

The program offers several advantages over conventional spreadsheets in purchase-order processing, searching, ledger reconciliation and vendor archiving. It is logically organized with various pop-up menus. Each screen looks like an empty form, containing boxes that need to be filled in. Once the form is completed, specific function keys or a simple carriage return execute the various tasks.

The purchase-order processing function allows entry of the order, with minimal repeat typing; if the item has been ordered before, then by checking the catalogue and marking the items, a new order can be generated and printed. The format of the order can be customized to print on any

can more easily have accurate figures for expendables. The program is divided into two parts — figures on the actual cost for each person, computed from their hourly or monthly rate of pay plus fringe benefits, and a break-down by grant of the percentage of that person's effort spent on each specific project. Printouts can be generated by grant to show the cost of a technician for each pay period or a summary of all personnel costs associated with a grant.

Both programs run on IBM-PC and compatible computers with a 512K RAM. I recommend the use of a computer with a 30–40 Mb hard drive, if the programs will be used to track a large number of grants or if the data need to be kept for a few years. *Grant Manager* is also available for the Apple Macintosh. □

Sharon Cohen is in the Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA.

## Fitting features

Bruno Orsi

**Enzfitter 1.05.** Biosoft, 22 Hills Road, Cambridge CB2 1JP/PO Box 580, Milltown, New Jersey 08850. £125, \$249.

TO THE inexperienced, a result of more than six figures printed on computer paper often assumes the status of an absolute truth. The judicious use of *Enzfitter* should provide such people, and others, with a more balanced view of the use of computers and statistics in the evaluation of experimental results. It is to the author's credit that he asserts that neither of these devices, for that is what they are, can "turn bad data into good".

*Enzfitter* is designed to be used on an IBM-PC (XT or AT) or a "true compatible". It is intended for curve-fitting generally, and more specifically for statistical analysis of ligand binding and for handling enzyme kinetic data. The program is versatile in that it will handle various types of screen graphics (CGA, EGA or Hercules) and a variety of printers (Epson compatible). It is pleasant to record the absence of a certain type of rodent as the author has decided to use the rather more effective method of a key input to initiate the desired course of action, the options being presented in a series of well-designed and simple overlapping 'pop-down' windows with textual prompts, rather than barely comprehensible icons.

The method of calculation uses a non-linear regression, even for linear equations, based on the Marquadt algorithm. It is rapid and does not depend on having good preliminary estimates of the constants. A notable feature of the program is the ability to show the fitted data in a screen graphics display. Optional numerical transformations are available, the most important of which are graphical displays of the residuals after the regression has been completed. Standard errors notwithstanding, this is surely the simplest and easiest way of assessing, literally at a glance, whether one has picked the right equation and whether the data are a good fit or not.

One of the most valuable features, that introduces a great deal of flexibility, is the capacity to input quite complex equations, or transformations of equations, of one's own choosing in an easy manner (a peculiarity of the version 1.05) I used, however, was its reluctance to accept  $x \Delta n$  ( $n = 1, 2, \dots$ ) for such equations, and it was necessary to use the rather cumbersome process of repeated multiplication). But my only real criticism of this software is in the treatment of weighting factors especially where both variables are subject to error, an extremely common

situation with ligand-binding data. Another improvement, useful for inhibition studies, would be the facility to enable the program to deal with a second independent variable.

Some typographical errors apart, the manual supplied with the software is more than adequate. It is clearly written, especially the chapter on the theoretical background which is well worth study in its own right. I do not entirely agree with the author's black-and-white attitude to the linear transformations of the Michaelis-Menten equation; this, perhaps, is reflected in his very selective

bibliography, in which he ignores a vast subculture that has been devoted to this problem for almost 60 years.

Is *Enzfitter* worth buying? The expert kineticist will almost certainly have his own software so this program is likely to be of passing interest only. But for the large number of research workers who want a good, rapid and visual analysis of their data, and who do not have the time or inclination to develop their own software, it is indeed a worthwhile purchase. □

Bruno Orsi is in the Department of Biochemistry, Trinity College, University of Dublin, Dublin 2, Ireland.

## Into solution

Gordon Reece

**MathCAD 2.0.** MathSoft, One Kendall Square, Cambridge, Massachusetts 02139/Freeport, Tamworth, Staffordshire B79 7BR. \$349, £345+VAT.

AS A BOY, embarking on what seems likely to be a lifetime of sums, I dreamed up the mathematician's equivalent of Sparky's Magic Piano. All you had to do was write down the sum to which you required an answer and the Magic Paper did all the work, setting everything out neatly. I have used *MathCAD* 2.0 for the past six weeks and it seems that my dream is pretty close to fulfilment.

*MathCAD* is "an editor that deals with equations" on an IBM-PC (or any compatible machine) with 512K RAM and CGA, EGA or Hercules graphics. The things it can do include all numerical calculations involving standard functions (including Bessel functions, statistical functions, real and complex fast Fourier transforms and their inverses). Integration is by Simpson's rule, differentiation by a four-point estimation process. Matrices are inverted and determinants evaluated by LU decomposition, using Crout's method with partial pivoting. Systems of equations and inequalities (called 'solve blocks') are solved by the Levenberg-Marquadt method. Full references are provided for all the algorithms used — the package is clearly designed by mathematicians for mathematicians. The algorithms work to specification and were extremely difficult to 'crash', despite my best efforts.

It took a day to come to terms with the editing facilities, using only the rather dry manual provided. It took me only another day to become hooked on the package. It loaded on a Toshiba T1000 laptop without the slightest difficulty, and the drivers provided for Epson dot-matrix and HP LaserJet printers worked perfectly. *MathCAD* occupied 360K of RAM but it recognized the LIM expanded memory.

At first I found the whole philosophy of *MathCAD* difficult, comparing the package unfavourably with Lotus 1-2-3 and on occasion seeing it as little more than an untidy spreadsheet. But this feeling evaporated as I began to make use of its potential.

Graphs can be plotted almost instantly. The quality is adequate for most purposes, but post-processing using a graphics package would normally be required for publication. The display and the printed output (which is just a copy of the screen) are not beautiful; they rely on the use of ASCII characters in full size throughout, including super- and subscripts, integral signs and square-root signs. IBM extended characters are available if the printer can produce them.

The algorithms are versatile and robust but occasionally rather slow. For example, to invert a  $7 \times 7$  matrix took 25 seconds, while Lotus 1-2-3 took less than one second. With a maths co-processor this problem would not be so serious. *MathCAD* balked at matrices larger than  $7 \times 7$  though the manual promised up to  $10 \times 10$ . There is some room for improvement here.

The package is almost ideal for generating work-sheets at all levels from primary-school to final-year undergraduate mathematics. It proved easy to design sheets of worked examples on Fourier series and Laplace transforms, vector and matrix operations. The possibilities for using it as a teaching and learning aid are mind-boggling. *MathCAD* does not provide facilities for drawing diagrams, nor are there 3D plotting routines — backs of envelopes will still be required. But these are minor shortcomings measured against the power of the package.

The price is currently rather high for individuals, but then the real price of a scientific calculator has fallen by 98 per cent in 15 years. If the same happens to *MathCAD* it, or something similar, will be an indispensable aid for every scientist — and every schoolchild — by 1995. □

Gordon Reece is in the Department of Engineering Mathematics, University of Bristol, University Walk, Bristol BS8 1TR, UK.



# GUIDE TO AUTHORS

*Please follow these guidelines so that your manuscript may be handled expeditiously.*

*Nature* is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully.

Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

## CATEGORIES OF PAPER

**Review Articles** survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

**Articles** are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

**Letters to Nature** are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

**Commentary** articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

**News and Views** articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

**Scientific Correspondence** is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

**Submission.** All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

## GENERAL

All contributions submitted for publication in *Nature* should conform with these rules.

**Manuscripts** should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

**Titles** should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

**Artwork** should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

**Colour Artwork.** A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

**Figure legends** must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

**References** should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unrefereed abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

**Abbreviations,** symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

**Footnotes** should not be used except for changed addresses or to identify the corresponding author if different from the first-named.

**Acknowledgements** should be brief. Grant numbers and contribution numbers are not allowed.

# The business of planet management

Euan G. Nisbet

*Decisions made about the world's economy have consequences for Earth's climate. At their forthcoming meeting in Toronto, the leaders of the West must take that into account.*

WHEN the leaders of seven Western nations meet next week in Toronto (19–21 June), they will be largely concerned with planning economic growth. But the scale of the global economy is now so great that they will implicitly also be planning the future of the Earth's atmosphere and of our climate. Should not the West's leaders look to the climatological consequences of their decisions?

The effects of human activities on the environment of our planet are much discussed but poorly understood. But we do know that the atmospheric concentrations of the greenhouse gases, especially CO<sub>2</sub>, CH<sub>4</sub> and the chlorofluorohydrocarbons (CFCs) are increasing, with the prospect that the radiative consequences by 2030 will be the equivalent of a doubling of CO<sub>2</sub> from pre-industrial concentrations.

Global warming may have been retarded by the thermal inertia of the oceans, but may now be perceptible. Political leaders at Toronto might note that some of the effects of global warming — a sharp reduction of soil moisture in mid-continental North America, for example — will have electoral implications. And who will be blamed for the consequences of a rise of sea level for coastal communities?

## Deforestation

The rapid deforestation of the tropics, especially Amazonia and the East Indies, is similarly alarming, because the forests are regulators of the atmospheric transfer of latent heat and moisture as well as processors of trace gases. Yet by planning for economic growth, the leaders at Toronto, unless they take thought, will be accelerating the rate of fossil fuel consumption and of deforestation. But to abate economic growth would be electoral suicide. Trapped, the Toronto leaders may defer the issue. They need to be offered alternatives.

Unless we are prepared to face the surprises of living in an atmosphere in which CO<sub>2</sub> has effectively been doubled, we have little choice but to abandon fossil fuel. Solar power and nuclear fusion are in the long run the most attractive alternatives, but neither can yet supplant fossil fuels, while hydropower has already flooded too much land. Only the nuclear-hydrogen economy offers an immediate escape: nuclear power can generate electricity to meet society's main need for

energy as well as for the production of hydrogen to fuel vehicles and remote machines. Properly and safely designed, such an economy would eliminate acid rain and allow CO<sub>2</sub> and CH<sub>4</sub> to fall to managed levels.

Nuclear power is deeply and deservedly unpopular, but the safety problems would be soluble if society would adopt a philosophy of passive safety rather than of maximum profit. The disposal of nuclear waste is a serious matter but, to a geologist, minor compared with the disposal of waste from coal. Perhaps the summit (and the rest of us) should consider how present patterns of energy use must seem to the developing countries. By continuing to burn oil and coal, thereby dispersing CO<sub>2</sub> around the globe, instead of facing the problems of handling nuclear power locally, the industrialized nations risk the reputation of playing fast and loose with the fates of poorer nations.

Hydrogen is the obvious currency of clean energy. Interestingly, in the late 1960s, Rhodesia (now Zimbabwe), with abundant electricity, seriously considered switching to a hydrogen economy only to discover that it lacked the industrial base for such an enterprise. But in the industrialized nations of the world, a switch from oil to hydrogen would be comparatively simple, would be profitable and would contribute to the long-term security of Western nations. Automobiles, after all, have an average life of a decade or less, and Detroit is used to periodic redesign and retooling. Without undue dislocation, the West's transport could be cleanly fuelled by 2000.

The forests present a different problem. Within 25 years, a large part of the forests of Amazonia and the East Indies will have gone. Much of the deforestation is the result of activities of large companies encouraged by government incentives (as in Borneo and Brazil, for example). Given government policies, the companies cannot be blamed. Western nations would do well to express their alarm in the strongest terms. They might also gently remind the Muslim nations of the East Indies of Muhammad's command to plant a tree "though the world end tomorrow".

But why not use the most powerful instrument for saving the tropical rainforests — third world debt? Most of the West's commercial banks wrote down the value of their loans to developing coun-

tries in 1987, but the debtors have so far been offered very little relief. In exchange for tax relief to the banks on their losses, the written-down debts could be transferred to central governments, which could then negotiate with developing countries (Brazil and Indonesia in particular) to cancel the debts in return for the preservation of the rainforests. Privately financed moves along the same lines should be encouraged. More generally, Western governments should consider buying debt from their banks and exchanging it for environmental protection in the tropics. The matter is urgent.

## Observatories

There is still more that can be done. The International Geosphere-Biosphere Programme (IGBP), which is the successor to the International Geophysical Year (IGY) and is now in its planning stages, includes among its proposals a scheme for a network of sophisticated Geosphere-Biosphere observatories. This is an enterprise whose cost is modest compared with the investments governments make in freeways, but which could assist enormously in the proper management of the Earth. IGBP deserves more than mere lukewarm approval at Toronto next week — it needs to receive active support and encouragement.

We are at a curious turning-point. The summit will be largely concerned with the global economy, which seems to be in reasonably good shape. At least in the prosperous nations, wealth continues to increase. But should not those at the meeting also pay attention to the long-term threats to the global economy represented by the prospect of climatic change? Is it not their duty, if only for economic reasons, to make a start on the proper management of the Earth?

But, the politicians will say, the scientific evidence for impending climatic change is not yet conclusive. That overlooks the truth that the dangers of inaction substantially outweigh the costs of action, not to mention the general conviction among scientists that the time has come to retreat from present policies. Is it too much to ask that the seven people meeting at Toronto next week should begin to see their problem (and ours) in that light? □

*Euan G. Nisbet is in the Department of Geological Sciences, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada.*



# Universal properties of dynamically complex systems: the organization of chaos

Itamar Procaccia

Department of Chemical Physics, The Weizmann Institute of Science, Rehovot, Israel 76100

*Natural systems far from equilibrium often show complex dynamic behaviour. There is fundamental progress in understanding universal aspects of the paths to such behaviour, of the trajectories just at the borderline of chaos and of the nature of the complexity in the chaotic region. A grammar of chaos, which sheds light on universal properties of complex systems, is emerging.*

THERMODYNAMICS is the appropriate framework for describing macroscopic natural systems whose intensive parameters (like temperature or pressure) are uniform in the bulk and on the system's boundaries. Under such uniform conditions a system is in equilibrium and the macroscopic variables are time independent. By subjecting the system to external forces (either at the boundaries or in the bulk), one pushes it out of equilibrium and may induce time dependence. Often, under the action of strong forces, macroscopic systems show very complex and turbulent behaviour. Common examples are waterfalls, storms and other flows of fluids under high pressure gradients. But examples are also found in other branches of physics, such as astrophysics, plasma physics and condensed matter physics, and there are counterparts in chemistry, biology and the social sciences as well.

The understanding, and even the precise characterization, of such fully developed turbulent behaviour is still an outstanding problem. Before reaching a highly turbulent condition, however, many systems pass through an intermediate state in which the time dependence of their state variables is already erratic but only a few modes are appreciably excited<sup>1</sup>. Such systems can be successfully described by mathematical models involving a finite, and even small, number of dynamic variables. It seems that in this regime of motion a surprising variety of *a priori* complex natural systems exhibit a complexity that is no worse than that captured by simple models involving a few variables.

Typically, the simple models involved in the theory of dynamically complex systems<sup>2</sup> are ordinary differential equations like

$$dx/dt = F_{\mu}(x) \quad (1)$$

where  $x$  is a point in a  $d$ -dimensional space of (the time-dependent) dynamic variables and  $\mu$  is a point in the space of (time-independent) parameters that can be typically controlled by the experimenter.  $F_{\mu}(x)$  is some nonlinear function of the dynamic variables. The dimension of the space of variables is referred to as the dimension of the dynamic system. By studying the dependence of the solutions of equations like (1) on the parameters we find regions of the parameter space  $\mu$  where the only available solutions are 'simple', that is, stationary solutions, periodic solutions and the like. In other regions of the  $\mu$ -space 'erratic' or 'chaotic' solutions also appear (defined below).

The borderline between the two regions of  $\mu$  is referred to as the 'borderline of chaos'<sup>3</sup>. Importantly, the sets of conditions characterizing the way that dynamical systems become chaotic as a result of following a path in the parameter space passing through the borderline of chaos are identical for widely different natural systems. It is most significant that some of these conditions enjoy universality in the sense that different systems display quantitatively identical behaviour along such paths in their parameter space (see below for examples).

The nature of the solution-trajectories  $x(t)$  of (1) at the borderline of chaos and in the chaotic regime is intricate and

complex. But again, at well-chosen points on the borderline of chaos, widely different systems display identical orbits—up to some irrelevant changes of coordinates: different systems have different dynamic variables, of course. Here I shall describe the ideas and methods needed to reach such a striking conclusion. Very recently the chaotic regime has also begun to yield to analysis and I shall discuss the exciting subject of the nature of complexity of chaotic orbits beyond the borderline of chaos. Recent progress indicates that complete enumerations of possible states is in principle possible and that this enumeration again enjoys universality<sup>4,5</sup>. This means there is hope for systematic modelling of dynamic systems in the intermediate regime of complexity before 'fully developed' turbulence.

It should be stressed that the natural systems exhibiting this intermediate regime of complexity are traditionally described by partial differential equations, such as the equations of fluid mechanics. Experimental evidence shows that a much more economical description with a few ordinary differential equations of the type of equation (1) are sufficient. As this fact is highly non-trivial and forms the basis for all the following treatment, I shall present the nature of this evidence and the reason for this behaviour immediately after defining chaotic behaviour.

## Low-dimensional chaos in natural systems

What is chaos? 'Chaotic' or 'erratic' behaviour has a very precise meaning in dynamics<sup>6</sup>. Two versions of the definition of chaos are most useful in natural applications, 'strong chaos' and 'weak chaos'. To understand these definitions and the nature of the dynamic complexity involved in so-called chaotic motion, consider the simple example furnished by the iteration scheme defined by

$$x_{n+1} = f(x_n) = 2x_n \bmod 1 \quad (2)$$

with initial conditions such that  $0 < x_0 < 1$ . Every initial condition  $x_0$  gives rise to an orbit  $x_0, x_1 = f(x_0), x_2 = f(f(x_0)), \dots$ . In this sense this dynamic system is deterministic and the future evolution of any orbit is known in principle. Such predictability does not exist in practice. Pick any nearby (in fact, *almost* any; do not pick rational  $x_0$ ) initial conditions  $x_0$  and  $x'_0$ , such that  $|x_0 - x'_0| = \Delta x$ . On iteration, the distance between the orbits,  $\Delta x_n = |x_n - x'_n|$  grows exponentially as  $2^n \Delta x_0$  until the two orbits lie on different sides of  $1/2$  in  $[0, 1]$ . This 'sensitivity to initial conditions' is the hallmark of strong chaos and destroys predictability. Uncertainty with respect to initial conditions always exists in natural systems and cannot be removed.

The weaker version of the definition of chaos has to do with the abundance of possible different dynamical asymptotic states, that is orbits, such as periodic orbits, that are allowed even after all transients have died out. This is shown in Fig. 1a where the function  $f(x)$  of equation (2) is plotted together with the function  $x_{n+1} = x_n$ . The common points  $x = 0$  and  $x = 1$  are unstable fixed

points of the iteration (2). Unstable here means that any deviation from the fixed point increases exponentially as a function of time. In Fig. 1b  $f(f(x)) = 4x \bmod 1$  is plotted again, together with  $x_{n+1} = x_n$ . Now there are four common points: the two original fixed points and two additional points belonging to an unstable periodic orbit of length 2 for the iteration (2). Similarly one can convince oneself that there are precisely  $2^n$  points belonging to unstable periodic orbits of length  $n$ .

This exponential growth in the number of longer-periodic orbits is the hallmark of weak chaos and is a necessary condition and prerequisite for strong chaos<sup>3,6</sup>. It turns out that there is a far better theoretical understanding of the onset of weak chaos than of strong chaos<sup>3</sup>. Mathematically it is easier to prove the abundance of asymptotic states than the actual existence of strong chaos. Moreover, it is becoming clear that the unstable periodic orbits organize the chaotic motion itself in a way that can be elucidated quantitatively<sup>4,5,7</sup> (see below). It should be stressed that the abundance of asymptotic states is only a necessary condition for strong chaos: if, as can happen, one of the available orbits is stable, chaotic motion is not observed. Chaotic motion can be seen when all the available orbits are unstable, and then can be conceived as a walk among the allowed motions. Even when an attracting orbit exists, the full complexity can be observed as a chaotic transient<sup>8-12</sup> that decays to stable periodic motion.

**Observation of low-dimensional behaviour.** The behaviour of natural systems can be modelled by low-dimensional dynamic systems (that is, systems like (1) with  $d$  small). A fluid flow is usually modelled with nonlinear partial differential equations like the Navier-Stokes equation

$$\frac{\partial \mathbf{v}}{\partial t} = -\mathbf{v} \cdot \nabla \mathbf{v} - \frac{1}{\rho} \nabla p + \frac{\eta}{\rho} \nabla^2 \mathbf{v} \quad (3)$$

where  $\mathbf{v}$ ,  $p$ ,  $\rho$  and  $\eta$  are the velocity field, pressure, density and viscosity respectively. With a bounded system one can expand the solution of (3) in a complete set of functions that satisfy the boundary conditions, say the set  $\phi_n(\mathbf{x})$ :

$$\mathbf{v}(\mathbf{x}, t) = \sum_{n=1}^{\infty} A_n(t) \phi_n(\mathbf{x}) \quad (4)$$

and upon substitution in (3), derive a set of ordinary nonlinear differential equations for the amplitudes

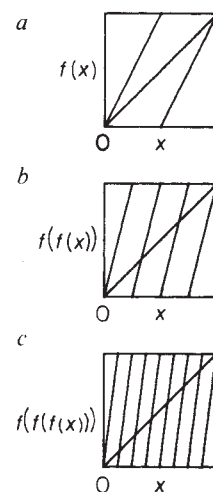
$$dA/dt = F(A) \quad (5)$$

However, this is an infinite dimensional dynamic system and such description seems unnecessarily cumbersome in view of the experimental evidence.

For concreteness let us consider two typical experiments that have recently become widespread<sup>13</sup>. The first is a thermoconvection experiment, in which a horizontal layer of fluid is heated from below (so-called Rayleigh-Bénard geometry)<sup>14</sup>, and in the second water in a cylindrical container is mounted on the cone of a loudspeaker so that the fluid oscillates precisely in the vertical direction. In the first case, the signals shown in Fig. 2a were obtained from a cell with an aspect ratio of 4, containing mercury, immersed in a magnetic field. The system is displaced from equilibrium by increasing the vertical temperature gradient, which in appropriate dimensionless units is called the Rayleigh number  $R$ . At a critical value of  $R$ ,  $R_c$ , the heat-conducting fluid at rest becomes unstable to convection, and convection rolls appear. At a value close to  $2R_c$  another instability sets in, and stationary convection disappears in favour of oscillatory convection, so that any variable measured at any point in space exhibits an oscillating signal. Figure 2a shows the succession of signals obtained upon increasing  $R$ —the now famous phenomenon of period doubling where the length of one period defined by a basic pattern repeated indefinitely, doubles at each successive instability.

The system responsible for these signals can be described by

**Fig. 1** a, Plot of the function  $f(x)$  of equation (1). The diagonal is the identity  $x_{n+1} = x_n$ , and the two common points of intersection of  $f(x)$  and  $x$  are fixed points of the iteration. b, Plot of the function  $f(f(x)) = 4x \bmod 1$ . Now there are four intersections with the identity  $x_{n+1} = x_n$ , two of which are identical to the fixed points in (a), and two new ones belonging to an unstable periodic orbit of length 2. c, Plot of the function  $(f(f(f(x)))) = 8x \bmod 1$ . Now there are eight intersections, and so on.



the equations of fluid mechanics. Consider, however, the succession of signals obtained from the simple one-dimensional transformation

$$x_{n+1} = g(x_n) = rx_n(1 - x_n) \quad (6)$$

in Fig. 2b. Although the signals in Fig. 2a become successively more complex, this is no worse than that shown in Fig. 2b. The low-dimensional model (6) is sufficient to describe the essential aspects of the dynamics. The correspondence between these two problems becomes more striking when we examine the parameter values ( $R_n$  and  $r_n$  respectively) where a period of length  $2^n \tau$  disappears in favour of a signal of length  $2^{n+1} \tau$ . It turns out that they form a geometric series with the same convergence rate:

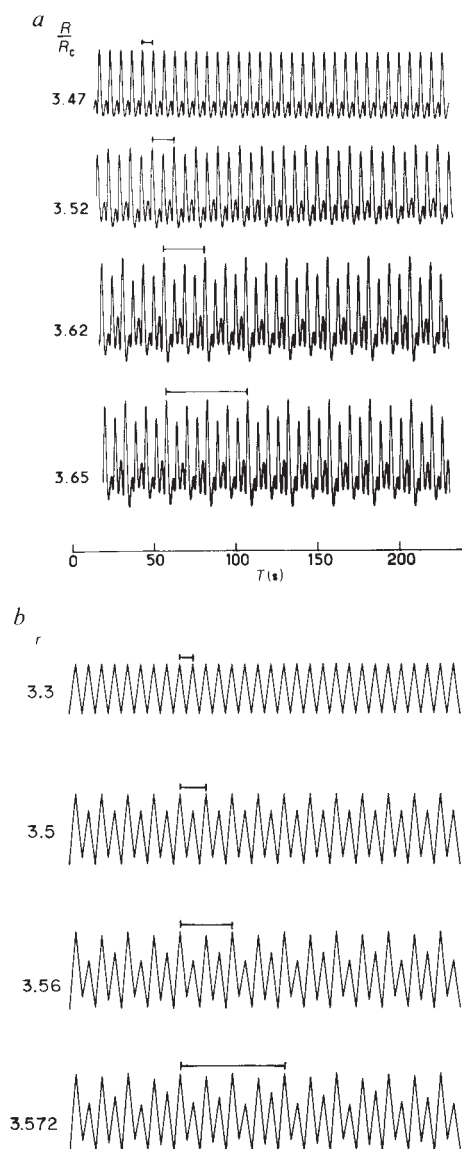
$$\delta = \lim_{n \rightarrow \infty} \frac{R_n - R_{n-1}}{R_{n+1} - R_n} = \lim_{n \rightarrow \infty} \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = 4.66902 \dots \quad (7)$$

This quantitative universality (here in parameter space) was one of the first major discoveries in the modern research of chaos<sup>15</sup>, and is quite well understood<sup>16,17</sup>. Below I shall stress more recently found aspects of universal behaviour, both on the borderline of chaos and in the chaotic regime itself.

Turning to the second example<sup>18</sup>, Faraday first observed that when the forcing in such an oscillating system exceeds some critical amplitude the surface of the water in the container deforms and surface waves appear. Using sophisticated optical and digitization techniques to follow the dynamics of the fluid surface, a typical orbit in the space of dynamic variables can be plotted (Fig. 3). This orbit looks rather complex, and divergence of nearby sections is quite apparent. It is, however, bounded in space and the question of what is the dimension of the geometrical object defined by the orbit asymptotically is important for revealing the minimal number of dynamic variables needed to describe the system as a function of time. The orbit is not one-dimensional, because it is not a curve closing upon itself, as it would be if the orbit were periodic. The dimension seems also to be smaller than three, because the orbit does not fill the three-dimensional volume.

If the dynamics of this system were described by an orbit in a space of 17 dynamic variables, projecting it onto two dimensions would result in a structureless blob and not in Fig. 3. In fact, the orbit defines asymptotically a 'fractal' set<sup>19</sup>, with a dimension between two and three. To determine the fractal dimension from the experimental data requires special algorithms<sup>20</sup>, which have been successfully applied to a variety of chaotic signals. The result in Fig. 3 is that the dimension is  $\sim 2.2$  (ref. 18). Beside the curious and interesting fact that the sets involved in chaos are fractal, the crucial point is that this dimension is small. Traditionally we tend to approach systems like this with partial differential equations but such an approach is enormously inefficient, as the data suggest that a low-





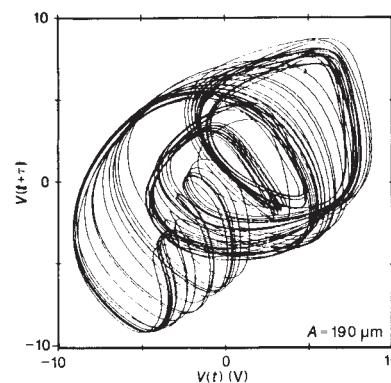
**Fig. 2** *a*, Period-doubling cascade in a thermoconvection experiment. The signal is the temperature of the fluid at one point as a function of time. The line segments indicate the length of one period, defined by a basic pattern repeated indefinitely, as a function of the control parameter  $R/R_c$ . Figure courtesy of A. Libchaber. *b*, Period-doubling cascade in the simple model (2). The orbits here are in discrete time, and orbit points were connected by straight lines to stress the quantitative similarity to (*a*). The values of  $r$  for which each orbit is obtained are indicated.

dimensional rather than an infinite-dimensional description should be sufficient.

**Why low dimensional?** To understand the huge reduction from infinite degrees of freedom to just a few, think again about the expansion of the fields that solve the partial differential equations of a given physical system in terms of a complete set of functions. In equilibrium, or in a stable stationary state, all these modes are stable in the sense that perturbations off the stationary state decay in time. Quite generally, we can write (5) as

$$\dot{\mathbf{A}} = \mathbf{L}\mathbf{A} + \mathbf{N}(\mathbf{A}) \quad (8)$$

where  $\mathbf{L}$  is a linear operator and  $\mathbf{N}$  is the nonlinear part. For small deviations from the steady state, the nonlinear part can be disregarded. Then stability means that all the eigenvalues  $\sigma_i$  of  $\mathbf{L}$  are lying in the left half of the complex plane, because any



**Fig. 3** A projection onto a two-dimensional plane of the orbit measured experimentally in a cylinder of fluid which exhibits surface waves as a result of a periodic oscillation. The variable measured,  $V$ , is related, through the experimental set-up, to the surface deformation. Rather than measure more than one variable, one measures time-delayed coordinates  $V(t)$ ,  $V(t+\tau)$ ,  $V(t+2\tau)$ , .... The orbit in a space of such time-delayed coordinates has the same invariant properties (such as dimension) as the orbit in the natural space of dynamical variables. The orbit can be shown to lie on a fractal object of dimension between two and three.

Figure courtesy of J. P. Gollub.

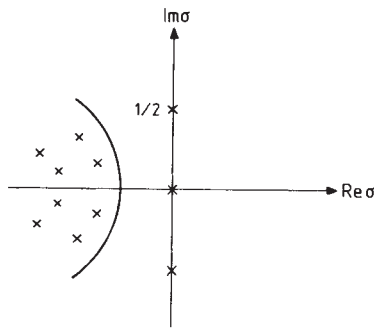
solution of a linear equation can be expanded in eigenmodes whose time dependence is  $e^{\sigma_i t}$ . (For an explanation of eigenvalues see the legend to Fig. 4.) A mode becomes unstable if it is associated with an eigenvalue that crosses the imaginary axis, gaining a positive real part that signals growth of perturbations in time. When a parameter is changed, at a critical value it often happens that only a few eigenvalues cross the imaginary axis, whereas all the infinity of other parameters are bounded away from the imaginary axis (Fig. 4). Whenever this situation occurs, the dynamics can be effectively reduced to a set of ordinary differential equations, where the number of variables involved is determined by how many modes cross, or have crossed, the imaginary axis<sup>21,22</sup>.

Intuitively, the reason for this reduction is simply separation of timescale. Modes that have just crossed the imaginary axis have a minute real part and are evolving on long timescales. All the other modes rapidly adapt themselves to the slow modes; they are 'enslaved' by them, technically known as the Centre Manifold Theorem, which characterizes the precise conditions for this to happen<sup>2</sup>.

## Metric universality at the borderline of chaos

Consider the experimental orbit shown in Fig. 5*a*, observed<sup>23,24</sup> at the borderline of chaos in a system that goes to chaos via the path called 'quasiperiodicity'. It was obtained from a rectangular cell filled with mercury, held under a temperature gradient to produce oscillatory convection, that is, the velocity field at any point in space is periodic in time. The system is immersed in a magnetic field with an alternating current sheet flowing through the conducting liquid. Owing to Lorentz force and Joule heating the current is well coupled to the hydrodynamic flow, and a quasiperiodic motion occurs, that is, a motion characterized by two frequencies, or an orbit which lies in an appropriate phase space on a torus. As the amplitude of the current increases, provided that the ratio of internal and forcing frequencies is kept as close as possible to an irrational number to avoid mode locking, the onset of chaos through quasiperiodicity is observed. At some threshold value, the torus on which the orbit lies wrinkles on all the length scales, the orbit becomes chaotic and the Fourier spectrum gains a broad-band component.

A major question, now affirmed, is whether this orbit (or any orbit at the borderline of chaos) possesses universal properties, such that different natural systems put in corresponding points of their parameter space would exhibit the same orbit up to a



**Fig. 4** Possible distribution of the eigenvalues of the linear operator  $L$  in the complex plane (see equation (8)). The eigenmodes of a linear operator are functions mapped by the operator to scalar multiples of themselves, and the eigenvalues are the scalars by which the functions are multiplied. Here one real and two complex-conjugate eigenvalues are crossing the imaginary axis simultaneously as a function of some control parameter. All the other eigenvalues are bounded away from the imaginary axis and their associated eigenmodes are rapidly decaying to equilibrium with the slow modes that form the basis for a low-dimensional description.

smooth change of coordinates. To answer this one has to decide what characteristics should be measured that remain invariant to a smooth change of coordinates. In principle, this information is supplied by the length scales of the orbit itself. Following the orbit in time, the distances between points become progressively smaller as more and more points are accumulated.

This detailed approach is at the heart of the theoretical development and leads to the so-called scaling function<sup>25,27</sup>, which is universal. For experimental purposes this approach is too fine and a more averaged or global approach is needed, such as the popular and successful one based on multifractal formalism<sup>28-30</sup>. In its simplest version, the space in which the orbit is embedded is partitioned into boxes of size  $l$  and the probability of falling in the  $i$ th box is counted, experimentally found as

$$P_i(l) = N_i(l)/N \quad (9)$$

where  $N_i(l)$  is the number of times the orbit falls in the  $i$ th box out of  $N$  points ( $N$  large) in the data set. The existence of such a probability is guaranteed by the existence of an ergodic natural measure<sup>6</sup> and in fact  $P_i(l)$  is just the integral over the measure in the box. There exist scaling exponents that describes the  $l$  dependence of  $P_i(l)$ ,

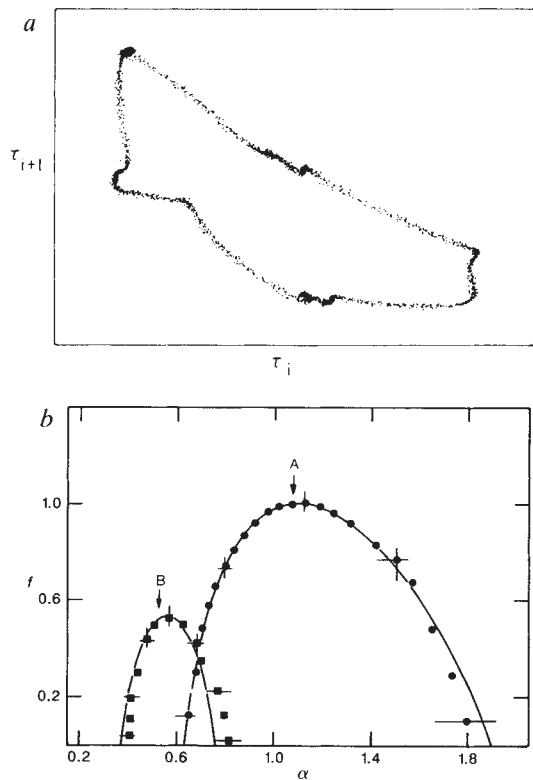
$$P_i(l) \propto l^\alpha \quad (10)$$

and sets at the borderline of chaos have a continuous range of scaling exponents  $\alpha$ ,  $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$  and that this range is universal (that is,  $\alpha_{\min}$  and  $\alpha_{\max}$  are universal). Also, one can see how typical a particular scaling exponent is by looking at the number of times  $N(\alpha)$  for which  $\alpha$  is found in a range between  $\alpha$  and  $\alpha + d\alpha$  as a function of  $l$ , where  $f(\alpha)$  is defined by

$$N(\alpha, l) \propto l^{-f(\alpha)} \quad (11)$$

$f(\alpha)$  can be interpreted as the dimension of a subset on which the scaling exponent is found to be  $\alpha$ . The natural measure is conceived as interwoven sets, each with different  $\alpha$ -values—hence the term ‘multifractal’: infinitely many fractal sets are bunched together. The orbit in Fig. 5a shows the variation in density that this approach attempts to quantify.

The importance of this approach lies in the fact that the function  $f(\alpha)$  can be shown theoretically to be universal for sets at the borderline of chaos. Methods have been advanced for the calculation of the  $f(\alpha)$  function from experimental signals<sup>24,26</sup> as well as from theoretical considerations<sup>26,27,31</sup>, and



**Fig. 5** *a*, The experimental orbit observed on a cut through a torus that is about to wrinkle at the onset of chaos resulting from quasiperiodicity. Note the variation in density. Theoretical analysis says that different natural systems at corresponding points of their parameter space have the same orbits up to a smooth change of coordinates. This set is topologically a circle, and its odd shape is typical of the arbitrary shapes that experimental trajectories take on. The ‘multifractal analysis’ extracts descriptors that are insensitive to coordinate changes. This figure has been obtained by strobing with the same frequency as one of the frequencies of the motion and plotting successive temperature readings as a function of each other, every time the orbit pierces through a plane that cuts orthogonally through the torus. *b*, The  $f(\alpha)$  function for the orbit of (a) (points of curve A) and the theoretical curve predicted *a priori* without any free parameters. The agreement is a demonstration of the metric universality of orbits of this kind. Curve B and the squares pertain to the  $f(\alpha)$  function at the borderline of chaos when the model is period doubling<sup>33</sup>.

comparisons between theory and experiment are possible. Universal numbers like  $\alpha_{\min}$ ,  $\alpha_{\max}$  and  $f_{\max}$  (the maximal point in the curve, which turns out to be the fractal dimension of the set) give an easily recognizable signature of the universal orbit in question. Figure 5b compares the  $f(\alpha)$  function of the orbit of Fig. 5a with the theoretical curve obtained without any adjustable parameters. An even more impressive agreement between the theoretical curve and an experimental function has been recently obtained<sup>22</sup> from quasiperiodic motion in a voltage-biased, cooled extrinsic p-type germanium. Similar results have been obtained in systems that reached the borderline of chaos by other paths, like period-doubling<sup>33</sup> (Fig. 5b).

Possibly the most important conclusion is that very simple models capture quantitatively the nature of orbits possessed by rather complex natural systems. The  $f(\alpha)$  function of Fig. 5b is correctly predicted by a simple model called the circle map, obtained from reducing the dynamics to that of the angle defined by the orbit on the cut torus,

$$\theta_{n+1} = \theta_n + \Omega - (K/2\pi) \sin 2\pi\theta_n \quad (12)$$

with appropriate values of  $K$  and  $\Omega$  which can be determined



*a priori*. Compared with the exceedingly difficult task, probably insurmountable with present-day knowledge, of solving the fluid-mechanical equations of the system responsible for the orbit in Fig. 5a, the model looks ridiculously simple. The power of universality is displayed here with flying colours.

### Universality in the chaotic regime

Because simple models are so successful in capturing universal properties at the borderline of chaos, are they also adequate for quantitative descriptions of chaotic motions? As mentioned before, there is no single allowed orbit in the chaotic regime that refines itself as time goes on. The dynamics becomes topologically complex in the sense that many, even infinitely many, different unstable periodic orbits are allowed. There are two problems: to describe the chaotic motion as the totality of the underlying available periodic orbits; and to organize the available periodic orbits with rules that allow *a priori* calculations on the basis of simple models. The second question can be answered in the affirmative only if the topology of the allowed motions is in some sense universal.

Recent theoretical progress indicates that the following statements can be made: (1) In some cases<sup>5</sup> chaotic systems have a finite codimension topological universality, that is, one can find points in parameter space ( $\mu$  of equation (1)) controlling a few parameters (say two) where every periodic orbit can be assigned a 'word' in an appropriate 'language' and a finite 'grammar' can be used to predict all the allowed words. Thus complete enumeration of the allowed motions is possible. (2) In other cases<sup>4</sup> the exact enumeration is a 'codimension infinity' problem: increasingly any parameters  $\mu$  need to be controlled to predict longer and longer periodic orbits. Here too one finds 'practical universality' because the problem can be set hierarchically, allowing an exponential increase in the length of predictable orbits with a linear increase in the number of parameters. (3) Control of all the allowed orbits provides exponentially good descriptions of the chaotic motion and the ergodic properties<sup>4,5,7</sup>.

Unlike the borderline problem, there is still no experimental corroboration of these statements: all that is known is theoretical. Further experimental and theoretical work is needed here.

These statements can be illustrated qualitatively with systems that go to chaos by way of quasiperiodicity. More precise definitions and notations are given in ref. 5. As mentioned before, upon entering the chaotic regime the 2-torus wrinkles on all the length scales and the simple model (12) can no longer be used. The motion is no longer on a smooth manifold, and the minimal model has to allow dynamics in the radial direction:

$$\begin{aligned}\theta_{n+1} &= \theta_n + \Omega - (K/2\pi) \sin 2\pi\theta_n + bg(r_n) \\ r_{n+1} &= bg(r_n) - (K/2\pi) \sin \pi\theta_n\end{aligned}\quad (13)$$

with  $g(r_n)$  being a function of  $r_n$ .

In the chaotic regime an infinite number of points exist in the parameter space, for which a topological theory can be adequately formulated. Figure 6 shows the shape of the chaotic orbit for one such point, through the dynamics of the angle  $\theta$ . The parameter values, say  $\tilde{K}$  and  $\tilde{\Omega}$ , are such that certain periodic orbits of length 2 and 3 are just becoming allowed. For  $K < \tilde{K}$  these orbits do not exist, whereas for  $K > \tilde{K}$  they are unstable. The shape of the set, with two turnbacks, suggests three regions of the allowed values of  $\theta$ , denoted  $L$ ,  $R$  and  $C$ .

We can now describe the 'language' of chaos. We can assign a 'word' to a periodic orbit length  $n$ ,  $\theta_1, \theta_2, \dots, \theta_n$ , writing  $L, R$  or  $C$  for every  $\theta_i$  that falls in one of these three regions. For example a period 3 can be  $(LRR)^\infty$  and a period 5  $(LCRLR)^\infty$ . A 'grammar' is a set of rules that determines which words are allowed. In the present case it turns out that all the words are composed of three primitive words, or blocks of symbols, denoted

$$\tilde{A} = LR, \quad \tilde{B} = LRR, \quad \tilde{C} = LCR \quad (14)$$

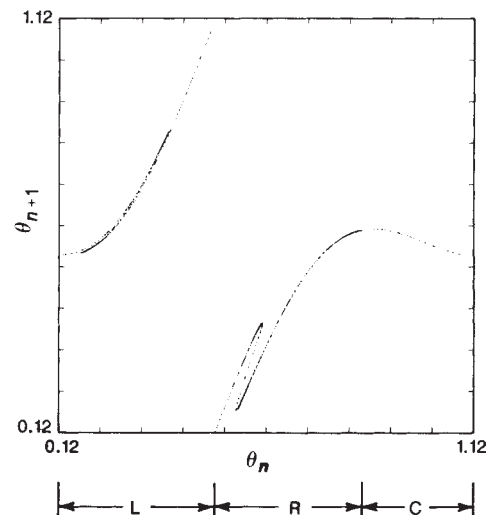


Fig. 6 An example of a chaotic attractor displayed by models like (13). Here  $g(r_n) = r_n$ ,  $b = 0.2$ ,  $K = 1.528$ ,  $\Omega = 0.58681$ . (Try it on your personal computer.) What is plotted is a 'return map' of  $\theta_{n+1}$  against  $\theta_n$ . The multisheeted nature (transverse Cantor set) is visible.

All possible combinations and permutations of  $\tilde{A}$ ,  $\tilde{B}$  and  $\tilde{C}$  are allowed, with the single constraint that  $\tilde{C}$  must be preceded by  $\tilde{A}$ . This grammar is universal and physical systems in corresponding points in their parameter space are therefore expected to exhibit periodic orbits that can be predicted *a priori* on the basis of simple models like (13). Exactly such a prediction is being tested now by J. Glazier at the Laboratory of A. Libchaber at the University of Chicago (Fig. 5).

Because the abundance of periodic orbits is a 'weak' definition of chaos, the set of all periodic orbits is not necessarily the observed attractor. When a chaotic attractor exists we can describe the motion itself as a random walk in the space of allowed unstable periodic orbits. A chaotic orbit defines an infinitely long word. Its description by periodic orbits is analogous to the description of real numbers (whose representation calls for an infinite string of digits which are the coefficients of an infinite sum of rational numbers). Ergodic properties of chaotic systems can be calculated as appropriately weighted averages on periodic motions.

By considering the set of all periodic points belonging to orbits of length  $n$ , and letting  $n$  increase, we achieve a hierarchical description of the chaotic dynamics. The indications are that this description converges very fast. The calculation of characteristics like the fractal dimension of the chaotic attractor<sup>34</sup> or the  $f(\alpha)$  function<sup>5,7</sup> of the invariant measure seem to converge well with this procedure.

The grammar of chaos seems to provide a useful way of describing the nature of universality in natural systems that are dynamically complex. Simple models may capture quantitative and detailed aspects of complex natural systems. Whether, and in what sense, this understanding can be used to improve the finite time predictions of the dynamics of chaotic systems is still an open question for the near future.

This work has been supported in part by the Israel Academy of Sciences, the Commission for Basic Research, and the Minerva Foundation, Munich, Germany. I thank P. Cvitanovic, J.-P. Eckmann, M. J. Feigenbaum, G. H. Gunaratne, M. H. Jensen, L. P. Kadanoff, A. Libchaber and C. Tresser for collaboration and discussion.

1. Ruelle, D. & Takens, F. *Commun. Math. Phys.* **20**, 167 (1971).

2. Guckenheimer, J. & Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1983).

3. Mackay, R. S. & Tresser, C. Warwick University preprint (1985).

4. Cvitanovic, P., Gunaratne, C. H. & Procaccia, I. *Phys. Rev. A* (in the press).
5. Gunaratne, G. H., Jensen, M. H. & Procaccia, I. *Nonlinearity* **1**, 157-180 (1988).
6. Eckmann, J. P. & Ruelle, D. *Rev. mod. Phys.* **57**, 617 (1985).
7. Gunaratne, G. H. & Procaccia, I. *Phys. Rev. Lett.* **54**, 1377 (1987).
8. Widom, M., Bensimon, D., Kadanoff, L. P. & Shenker, S. F. *J. statist. Phys.* **32**, 443 (1983).
9. Kadanoff, L. P. & Tang, C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1276 (1984).
10. Kantz, H. & Grassberger, P. *Physica D* **17**, 75 (1985).
11. Szepfalussy, P. & Tel, T. *Phys. Rev. A* **34**, 2520 (1986).
12. Tel, T. *Phys. Rev. A* **36**, 1502 (1987).
13. Berge, P., Pomeau, Y. & Vidal, C. *Order within Chaos* (John Wiley, New York; Herman, Paris, 1984).
14. Libchaber, A., Fauve, S. & Laroche, C. *Physica D* **7**, 73 (1983).
15. Feigenbaum, M. J. *J. statist. Phys.* **19**, 25 (1978); **21**, 669 (1979).
16. Cvitanovic, P. *Universality in Chaos* (Adam Hilger, Bristol, 1984).
17. Hao-Bai-Lin *Chaos* (World Scientific, Singapore, 1984).
18. Ciliberto, S. & Gollub, J. P. *Phys. Rev. Lett.* **52**, 922 (1984); *J. Fluid Mech.* **158**, 381 (1985).
19. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).
20. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346 (1983); *Physica D* **9**, 189 (1983).
21. Couillet, P. & Spiegel, E. A. *SIAM J. appl. Math.* **43**, 776 (1983).
22. Meron, E. & Procaccia, I. *Phys. Rev. Lett.* **56**, 1323 (1986); *Phys. Rev. A* **34**, 3221 (1986); *Phys. Rev. A* **35**, 4008 (1987).
23. Stavans, J., Heslot, F. & Libchaber, A. *Phys. Rev. Lett.* **55**, 596 (1985).
24. Jensen, M. H., Kadanoff, L. P., Libchaber, A., Procaccia, I. & Stavans, J. *Phys. Rev. Lett.* **25**, 2798 (1985).
25. Feigenbaum, M. J. *Commun. Math. Phys.* **77**, 65 (1980).
26. Feigenbaum, M. J., Jensen, M. H. & Procaccia, I. *Phys. Rev. Lett.* **57** (1986).
27. Jensen, M. H., Kadanoff, L. P. & Procaccia, I. *Phys. Rev. A* **36**, 1404 (1987).
28. Hentschel, H. G. E. & Procaccia, I. *Physica D* **8**, 435 (1983).
29. Frisch, U. & Parisi, G. in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* (ed. Ghil, M., Benzi, R. & Parisi, G. 84 (North-Holland, Amsterdam, 1985).
30. Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. & Shraiman, B. I. *Phys. Rev. A* **33**, 1141 (1986).
31. Feigenbaum, M. J. *J. Statist. Phys.* **46**, 919 (1987); **46**, 925 (1987).
32. Gwinn, E. G. & Westervelt, R. M. *Phys. Rev. Lett.* **59**, 157 (1987).
33. Glazier, J. A., Jensen, M. H., Libchaber, A. & Stavans, J. *Phys. Rev. A* **34**, 1621 (1986).
34. Auerbach, D., Cvitanovic, P., Eckmann, J.-P., Gunaratne, G. H. & Procaccia, I. *Phys. Rev. Lett.* **58**, 2387 (1987).

## ARTICLES

# Isotope evidence of a mantle convection boundary at the Australian-Antarctic Discordance

E. M. Klein\*, C. H. Langmuir\*, A. Zindler\*, H. Staudigel† & B. Hamelin\*‡

\* Lamont-Doherty Geological Observatory and Department of Geological Sciences of Columbia University, Palisades, New York 10964, USA

† Scripps Institution of Oceanography, La Jolla, California 92093, USA

‡ Laboratoire de Géochimie et Cosmochimie, 75230 Paris, France

*Basalts from the Southeast Indian Ridge south of Australia form two geographically and isotopically distinct groups that show affinities with either Indian Ocean or Pacific/Atlantic Ocean isotope compositions. The data raise the possibility that there is a sharp boundary between the Indian and Pacific Ocean isotope provinces. The isotope boundary occurs over less than 200 km within the Australian-Antarctic Discordance, a tectonically complex region believed to overlie a zone of downwelling mantle flow.*

THE distribution of chemical heterogeneities in the mantle and the relationship of these heterogeneities to surface features of the Earth are fundamental problems in studies of mantle dynamics. A more complete understanding of the composition and physical organization of mantle heterogeneities would place constraints on models of solid-earth geochemical cycles and the kinematics of mantle convection. Our knowledge of the identities and distribution of distinct mantle reservoirs is largely inferred from the manifestation of three-dimensional mantle variability in the two-dimensional oceanic crust. It has recently become clear that there are regionally distinct and mappable domains in the isotope composition of the oceanic crust, and the geometry of these domains can be used to characterize the distribution of sub-oceanic mantle compositions and to infer patterns of mantle flow.

One important discovery is that mid-ocean ridge basalts (MORB) from the Indian Ocean have isotope compositions that are distinct from those of Atlantic and Pacific MORB<sup>1-12</sup>. As the source of normal MORB is probably in the upper mantle, the existence of this large-scale, mappable, isotope unit implies that there are distinct upper mantle convection regimes that have not been homogenized by convective stirring. Little is known, however, about the location of the boundaries between these isotope domains, and how these domains and their boundaries relate to mantle convection and surficial tectonic features.

Here we present Pb, Sr and Nd isotope compositions of basalts collected along the Southeast Indian Ridge between Australia and Antarctica, and suggest that there is a remarkably abrupt eastern boundary to the Indian Ocean geochemical province. The apparent isotope boundary is located within the

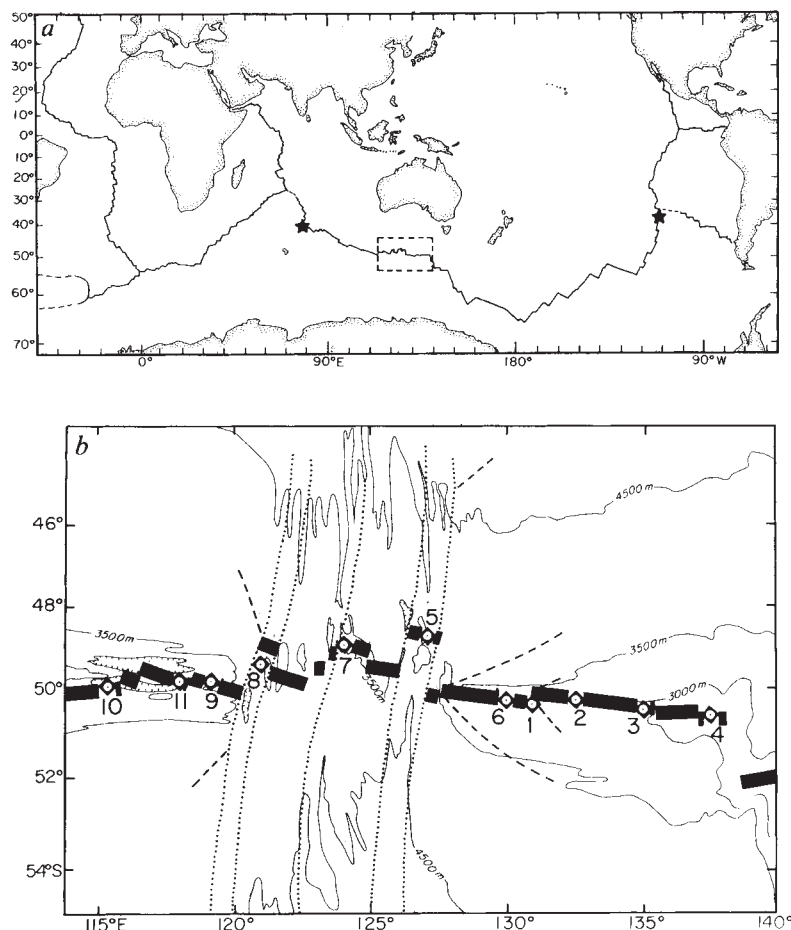
anomalously deep and tectonically complex Australian-Antarctic Discordance<sup>13-15</sup>. We then explore the significance of this boundary for models of mantle convection.

## Geographic setting and sample location

The Southeast Indian Ridge (SEIR) separates the Indo-Australian and Antarctic plates (Fig. 1a). It extends south-east from the Indian Ocean triple junction in the central Indian Ocean, trends east-west between Australia and Antarctica, and terminates at the Macquarie triple junction in the south-west Pacific Ocean. In the region between Australia and Antarctica, both the ridge axis and its flanks are anomalously deep compared with the empirical age versus depth relation seen in other oceans<sup>13,16</sup>. The ridge axis reaches a depth of >4,000 m within a 500 km-wide segment of the ocean basin called the Australian-Antarctic Discordance (AAD; Figs 1b, 2a). The AAD is characterized by rough topography, low magnetic anomaly amplitudes, and a high density of north-south trending fracture zones<sup>13-18</sup>. Based on morphological and geophysical data, the SEIR south of Australia has been divided into three zones<sup>14</sup>: the ridge segments east and west of the AAD have been respectively designated as zone A and zone C; and the AAD itself as zone B (here called the AAD).

Eleven dredge hauls were made along a 1,700 km-long segment of the SEIR between 115°E and 138°E during the austral summer of 1975-76 aboard the R/V *Vema*<sup>18</sup> (Fig. 1b). Three dredge hauls were made within the AAD, five to the east within zone A, and three to the west within zone C (Figs 1b, 2a). Studies on the major-element chemistry of basalts recovered from these dredges were performed to explore the relationship





**Fig. 1** *a*, Global ocean ridge system. Dashed box encompasses the area of study shown in panel *b*. Stars show locations of nearest other ridge axis samples for which isotope data are currently available<sup>7,44</sup>. *b*, General bathymetry and tectonic elements of the Southeast Indian Ridge, after Hayes and Conolly<sup>13</sup> and Vogt *et al.*<sup>17</sup>; 3,000 m, 3,500 m, and 4,500 m bathymetric contours are shown. Also shown are ridge segments (heavy lines), dredge locations (open diamonds), fracture zones (dotted lines) and propagating rift pseudofaults (dashed lines).

between chemistry and magnetic anomaly amplitudes. We have undertaken a more extensive study of the major, trace and rare-earth element and isotope composition of samples from these dredges (ref. 19; E.M.K., C.H.L. and H.S., manuscript in preparation) and some of these results have been examined in a global context by Klein and Langmuir<sup>20</sup>.

Anderson *et al.*<sup>18</sup> reported several Sr isotope analyses of whole-rock powders from these dredge hauls, but their highly radiogenic values, which approach or exceed those of sea water, suggest that these analyses suffer from the contamination problems common in whole-rock analyses of Sr isotopes. In addition, Cohen *et al.*<sup>21</sup> reported Pb, Sr and Nd isotope analyses of one basaltic glass from dredge D1; Hamelin *et al.*<sup>7</sup> described the isotope composition of this sample as that of standard MORB, contrasting it with the Indian Ocean signature. Thus, the one high-quality isotope analysis previously reported on this suite of dredge hauls suggested that the region is distinct from the Indian Ocean isotope province.

## Methods

All isotope analyses were performed on natural basaltic glasses, with the exception of the Pb analysis for sample D6-1, which was performed on fresh whole-rock powder because of the paucity of glass. Glass chips were hand-picked under a binocular microscope to avoid all visible signs of alteration and to minimize the inclusion of phenocrysts. To remove possible surface contamination, samples were mildly leached for 20 min in 1 N HCl at room temperature in an ultrasonic bath, and then ultrasonically cleaned in distilled water before dissolution in HF.

For Nd, Sm, Sr and Rb analyses, ~100 mg of sample was spiked with a mixture of enriched isotopes of <sup>150</sup>Nd, <sup>149</sup>Sm, <sup>84</sup>Sr and <sup>87</sup>Rb. Chemical separation and analytical techniques used for analysis of these elements have been described by Zindler

*et al.*<sup>22</sup>. For Pb analyses, ~300 mg of sample was dissolved and chemical separation was performed using a technique patterned after that described by Manhès *et al.*<sup>23</sup>. Pb separates were loaded in silica gel and H<sub>3</sub>PO<sub>4</sub> onto single, zone-refined Re filaments and data were taken at several temperatures between 1,100 °C and 1,200 °C. All measurements were performed on Lamont's modified V.G. Micromass 30 mass spectrometer<sup>22</sup>.

## Results

Pb, Sr and Nd isotope compositions of SEIR samples analysed in this study are presented in Table 1 and Figs 2–4. The data display substantial variations in all three systems. The SEIR data span, for example, more than half of the range in Nd–Sr isotope compositions found in MORB from the Atlantic, Pacific and Indian Oceans (Fig. 4).

In terms of along-strike systematics (Fig. 2), the data show a general trend of decreasing <sup>206</sup>Pb/<sup>204</sup>Pb from east to west, although D7, located in the centre of the AAD, shows lower <sup>206</sup>Pb/<sup>204</sup>Pb than both of its adjacent dredges. Sr and Nd isotopes show no gradual trends from east to west; samples from all dredge hauls east of and including D5 show lower Sr and higher Nd isotopic ratios compared with all dredges west of and including D7.

In Figs 3 and 4, our SEIR data are compared with fields for data obtained on MORB from the North Atlantic, Pacific and Indian Oceans. All samples west of and including D7 (dredges 7–11) fall within the distinct field for Indian Ocean MORB on plots of <sup>206</sup>Pb/<sup>204</sup>Pb versus <sup>208</sup>Pb/<sup>204</sup>Pb or <sup>143</sup>Nd/<sup>144</sup>Nd. On a plot of <sup>206</sup>Pb/<sup>204</sup>Pb versus <sup>87</sup>Sr/<sup>86</sup>Sr, all but one sample from the western group also plot entirely within the Indian Ocean MORB field, which is characterized by high <sup>87</sup>Sr/<sup>86</sup>Sr at low <sup>206</sup>Pb/<sup>204</sup>Pb. The exceptional sample, from dredge D10, displays even more radiogenic Sr at low values of <sup>206</sup>Pb/<sup>204</sup>Pb than the

**Table 1** Compositions of Southeast Indian Ridge basalts

| Dredge<br>-sample | Location                   | Dredge<br>depth<br>(m) | <sup>206</sup> Pb/ <sup>204</sup> Pb | <sup>207</sup> Pb/ <sup>204</sup> Pb | <sup>208</sup> Pb/ <sup>204</sup> Pb | <sup>87</sup> Sr/ <sup>86</sup> Sr | <sup>143</sup> Nd/ <sup>144</sup> Nd | Sr         | Rb   | Nd            | Sm   |
|-------------------|----------------------------|------------------------|--------------------------------------|--------------------------------------|--------------------------------------|------------------------------------|--------------------------------------|------------|------|---------------|------|
| D4-1              | 50°12.9' S<br>137°33.3' E  | 2,549                  | 18.812<br>18.820                     | 15.503<br>15.506                     | 38.166<br>38.144                     | 0.70253 ± 3                        | 0.513139 ± 23                        | 107        | 0.32 | 9.44          | 3.39 |
| D3-4              | 50°25.0' S<br>135°05.4' E  | 2,910                  | 18.998<br>18.959                     | 15.613<br>15.566                     | 38.468<br>38.363                     | 0.70257 ± 4                        | 0.513074 ± 19                        | 178        | 0.59 | 12.20         | 3.80 |
| D2-19             | 50°16.0' S<br>132° 33.0' E | 3,215                  | 18.911                               | 15.514                               | 38.319                               | 0.70261 ± 4                        | 0.513058 ± 14                        | 115        | 0.50 | 13.15         | 4.55 |
| D1-2              | 50°24.5' S<br>131° 00.3' E | 3,091                  | 18.805                               | 15.499                               | 38.262                               | 0.70264 ± 2                        | 0.513123 ± 15                        | 233        | 2.62 | 16.01         | 4.64 |
| D6-1              | 50°18.0' S<br>130°02.5' E  | 3,106                  | 18.617                               | 15.480                               | 38.095                               | 0.70259 ± 2                        | 0.513102 ± 17                        | 115        | 0.88 | 13.99         | 4.80 |
| D5-5              | 48°44.3' S<br>127°04.8' E  | 4,099                  | 18.572                               | 15.482                               | 38.097                               | 0.70255 ± 3                        | 0.513118 ± 24<br>*0.513108 ± 21      | 131        | 0.52 | 8.58<br>*8.60 | 3.00 |
| D7-3              | 49°02.0' S<br>124°00.0' E  | 3,989                  | 18.057                               | 15.439                               | 37.858                               | 0.70290 ± 3                        | 0.513037 ± 17                        | 179        | 3.92 | 14.48         | 4.29 |
| D7-7              | 49°02.0' S<br>124°00.0' E  | 3,989                  | 18.008                               | 15.462                               | 37.794                               | 0.70300 ± 3                        | 0.513007 ± 21                        | 148        | 1.28 | 7.66          | 2.54 |
| D8-8              | 49°28.0' S<br>121°02.0' E  | 3,532                  | 18.225<br>18.220                     | 15.465<br>15.464                     | 38.211<br>38.212                     | 0.70314 ± 2                        | 0.512972 ± 10                        | 177        | 6.10 | 10.76         | 3.14 |
| D9-2              | 49°48.5' S<br>119° 10.5' E | 3,323                  | 18.248                               | 15.489                               | 38.003                               | 0.70293 ± 3                        | 0.513036 ± 20                        | 124        | 1.19 | 9.06          | 3.08 |
| D11-6             | 49°51.5' S<br>118°00.0' E  | 3,733                  | 17.944                               | 15.409                               | 37.743                               | 0.70283 ± 3<br>0.70284 ± 4         | 0.513041 ± 21                        | 145<br>148 | 0.75 | 7.49          | 2.54 |
| D10-10            | 49°55.0' S<br>115°22.5' E  | 3,087                  | 17.772<br>17.756                     | 15.494<br>15.472                     | 37.837<br>37.768                     | 0.70346 ± 4<br>0.70344 ± 2         | 0.512978 ± 14<br>0.512997 ± 20       | 143        | 1.08 | 7.72          | 2.55 |

<sup>87</sup>Sr/<sup>86</sup>Sr ratios are normalized to a value of 0.11940 for <sup>86</sup>Sr/<sup>88</sup>Sr and are reported relative to a value of 0.70800 for <sup>87</sup>Sr/<sup>86</sup>Sr for the Eimer and Amend SrCO<sub>3</sub> standard. <sup>143</sup>Nd/<sup>144</sup>Nd ratios are normalized to a value of 0.72190 for <sup>146</sup>Nd/<sup>144</sup>Nd and are reported relative to a value of 0.51265 for <sup>143</sup>Nd/<sup>144</sup>Nd for BCR-1. Pb data are normalized to the values for SRM 981 reported by Hamelin *et al.*<sup>42</sup> (16.937, 15.491 and 36.704 for <sup>206</sup>Pb/<sup>204</sup>Pb, <sup>207</sup>Pb/<sup>204</sup>Pb and <sup>208</sup>Pb/<sup>204</sup>Pb, respectively), which for <sup>208</sup>Pb/<sup>204</sup>Pb differs slightly from the value reported by Catanzaro *et al.*<sup>43</sup>. Mean measured values of SRM 981 are 16.879 ± 0.038, 15.426 ± 0.040 and 36.512 ± 0.117 for <sup>206</sup>Pb/<sup>204</sup>Pb, <sup>207</sup>Pb/<sup>204</sup>Pb and <sup>208</sup>Pb/<sup>204</sup>Pb, respectively (2σ; 11 runs). All reported duplicate measurements were made on separate dissolutions of the sample, with the exception of the duplicate marked by an asterisk, which was made on the same dissolution. Duplicate measurements for Pb isotope determinations agree to better than 0.05%/AMU, with the exception of the measurements for D3-4, which agree to better than 0.1%/AMU. Analytical uncertainties for Sr and Nd isotope ratios are 2σ/√N. Blank values for Sr, Nd and Pb average 0.45, 0.07 and 0.26 ng respectively and are below the level of significance. Pb and Sr isotope analyses of D1-2 are identical within analytical uncertainty to analyses of a different sample from this dredge haul reported by Cohen *et al.*<sup>21</sup>, these authors report a lower value of 0.51305 for <sup>143</sup>Nd/<sup>144</sup>Nd (the reason for this difference is unclear).

Indian Ocean MORB field shown in Fig. 3 (note that the high <sup>87</sup>Sr/<sup>86</sup>Sr value for this sample has been confirmed by a replicate analysis of a separate sample aliquot). In contrast, all samples east of and including D5 (dredges 1–6) fall within or close to the fields for North Atlantic and Pacific ridges on plots of <sup>206</sup>Pb/<sup>204</sup>Pb versus <sup>208</sup>Pb/<sup>204</sup>Pb, <sup>87</sup>Sr/<sup>86</sup>Sr or <sup>143</sup>Nd/<sup>144</sup>Nd (Fig. 3). On a plot of <sup>206</sup>Pb/<sup>204</sup>Pb versus <sup>207</sup>Pb/<sup>204</sup>Pb, these samples, with the exception of D3, also form a linear array that is characteristic of North Atlantic/Pacific MORB, but they are offset to lower values of <sup>207</sup>Pb/<sup>204</sup>Pb.

Thus, the SEIR data form two distinct, non-overlapping isotopic groups that correspond to longitudinal position along the ridge axis: an eastern group and a western group. Samples from the western group show strong isotopic affinities with Indian Ocean MORB, whereas samples from the eastern group show isotopic affinities with Pacific and North Atlantic MORB. The eastern group is composed of all dredges within zone A (dredges 1, 2, 3, 4 and 6) and the adjacent easternmost dredge within the AAD (dredge 5); the western group is composed of dredges 7 and 8 within the AAD and all dredges within zone C (dredges 9, 10 and 11). It should be emphasized that no sample with Pb isotope composition like any eastern group sample has yet been observed in the Indian Ocean, and conversely, no sample with Pb, Sr and Nd systematics like any western group sample has yet been observed along either the Pacific or North Atlantic Ocean ridge system.

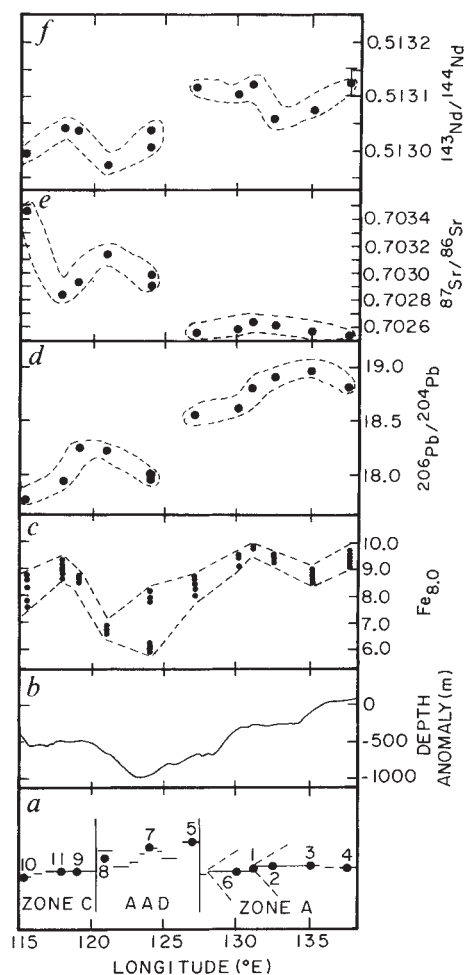
Subtle but important differences have also been noted between the isotopic compositions of North Atlantic and Pacific MORB. White *et al.*<sup>10</sup> and Ito *et al.*<sup>11</sup> have observed that Pacific MORB

show more uniform <sup>87</sup>Sr/<sup>86</sup>Sr ratios and lower <sup>143</sup>Nd/<sup>144</sup>Nd for the same <sup>87</sup>Sr/<sup>86</sup>Sr, compared with Atlantic MORB. Indeed, samples from the eastern group cluster on a plot of <sup>87</sup>Sr/<sup>86</sup>Sr versus <sup>143</sup>Nd/<sup>144</sup>Nd (Fig. 4) and extend to lower values of <sup>143</sup>Nd/<sup>144</sup>Nd for the same <sup>87</sup>Sr/<sup>86</sup>Sr than observed in most North Atlantic MORB. Samples from the eastern group, however, also extend to higher values of <sup>206</sup>Pb/<sup>204</sup>Pb than N-MORB from the Pacific (Fig. 3), and do not plot entirely within the Pacific fields in Fig. 3a–c. Inclusion of more enriched samples from the ridges associated with the Easter microplate would expand the Pacific MORB field to these, more radiogenic values<sup>10</sup>. Thus, although the Pb isotopic compositions for the eastern group samples appear to be somewhat 'enriched', their Nd–Sr systematics suggest a greater isotopic affinity with Pacific MORB than with North Atlantic MORB.

## Discussion

The distinction in isotope composition between the western and eastern groups, and their isotope affinities with either Indian Ocean or Pacific and North Atlantic Ocean isotope compositions raise the possibility that this region forms a boundary between the Indian and Pacific Ocean isotope provinces. This hypothesis should eventually be tested by sampling of the vast unsampled or sparsely sampled segments of the adjacent SEIR and Pacific–Antarctic ridges (Fig. 1a). Indeed, further sampling of the Pacific–Antarctic ridge may reveal that the more radiogenic lead ratios seen in the eastern group from the SEIR are typical of the southern or eastern Pacific Ocean ridge system. The current evidence of a boundary between the Indian Ocean and Pacific Ocean isotope provinces, however, is strong enough to warrant



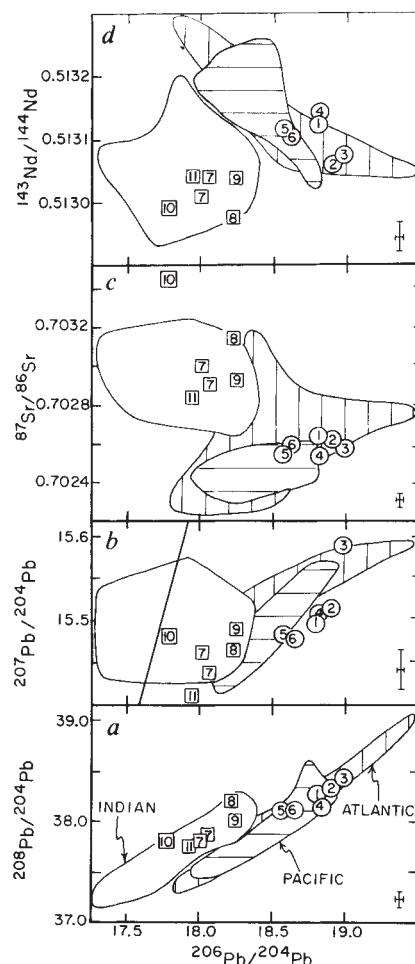


**Fig. 2** Longitude ( $^{\circ}$ E) versus *a*, tectonic features along the SEIR, including fracture zones bounding the AAD, propagating rifts (dashed) and dredge locations (after refs 17 and 18); *b*, depth anomalies along the ridge axis (after ref. 18); *c*, mean FeO contents of 2–13 basalts from each dredge, normalized to 8%  $\text{MgO}^{19,20}$ ; *d*,  $^{206}\text{Pb}/^{204}\text{Pb}$ ; *e*,  $^{87}\text{Sr}/^{86}\text{Sr}$ ; *f*,  $^{143}\text{Nd}/^{144}\text{Nd}$ . Typical analytical uncertainties ( $\pm 2\sigma/\sqrt{N}$ ) for  $^{143}\text{Nd}/^{144}\text{Nd}$  are shown by error bars on one data point; uncertainties for  $^{206}\text{Pb}/^{204}\text{Pb}$  and  $^{87}\text{Sr}/^{86}\text{Sr}$  are less than or equal to the size of the data points.

investigation of the implications of such a boundary for mantle composition and convection.

**The boundary.** Our data suggest that the geographical boundary between the Indian Ocean and Pacific Ocean isotope provinces occurs between D5, which shows Pacific-like isotope compositions, and D7, which shows Indian Ocean-like compositions (Figs 2–4). Between these two dredges the ridge axis is truncated by one large-offset transform (100 km) and at least one and possibly two small-offset ( $<30$  km) transforms<sup>17</sup> (Fig. 1b). The unsampled ridge segments between these two dredges total  $\sim 200$  km in length and encompass much of the eastern half of the AAD. It must be emphasized that in no place other than the AAD has a fundamental change from one MORB isotope province to another been observed. In a global context, it is thus remarkable that this change appears to occur over less than 200 km of ridge length. In the absence of sampling between dredges 5 and 7, it is not known whether this apparent isotope boundary is gradational over the intervening ridge length or occurs abruptly.

To understand the physical basis of the isotopic boundary within the AAD, it is useful to examine the geophysical information that exists for this region. One of the most striking features



**Fig. 3**  $^{206}\text{Pb}/^{204}\text{Pb}$  versus *a*,  $^{208}\text{Pb}/^{204}\text{Pb}$ ; *b*,  $^{207}\text{Pb}/^{204}\text{Pb}$ ; *c*,  $^{87}\text{Sr}/^{86}\text{Sr}$ ; and *d*,  $^{143}\text{Nd}/^{144}\text{Nd}$ . Data are plotted by dredge number: 'eastern group' sample numbers are enclosed by circles; 'western group' sample numbers are enclosed by squares. Analytical uncertainties are shown by crosses in each panel. Geochron in *b* is calculated for 4.55 Gyr, assuming initial Pb isotope composition of Canyon Diablo<sup>45</sup>. Fields for N-MORB isotope data from North Atlantic (vertical ruled), Pacific (horizontal ruled), and Indian (open) Ocean ridges are shown for comparison; data for MORB fields are from refs 3–11, 21, 44, 46–48, renormalized to a comparable basis. Excluded from these N-MORB fields are samples that show the bathymetric or chemical influence of nearby hot spots (defined by criteria described by Ito *et al.*<sup>11</sup>). Three Indian Ocean samples with  $^{87}\text{Sr}/^{86}\text{Sr} > 0.7045$  are omitted from the diagrams because they would substantially compress the  $^{87}\text{Sr}/^{86}\text{Sr}$  scale<sup>6,7</sup>. The two samples recovered from the Sheba ridge near the Gulf of Aden<sup>4,21</sup> are also not included because they may be isotopically distinct from other Indian Ocean ridge volcanics<sup>12</sup>.

of this segment of the SEIR and its flanks is its anomalous depth<sup>13,16</sup>. On the ridge axis, the depth anomaly reaches a maximum of 1 km greater than average ridge depths in the centre of the AAD at  $\sim 124^{\circ}$  E (the location of D7, the easternmost dredge showing Indian Ocean-like chemistry; Fig. 2). In addition, a recent analysis of Seasat altimeter data shows a pronounced, saddle-shaped, negative geoid anomaly centred on the ridge axis at  $\sim 125^{\circ}$  E (ref. 24), and satellite gravimetry measurements show a broad, negative free-air anomaly centred on the AAD (ref. 16). Based primarily on the anomalous depth within the AAD and the associated shoaling to its east and west, Hayes and Conolly<sup>13</sup> and Weissel and Hayes<sup>16</sup> suggested that the unique bathymetry may result from downwelling convective flow within the asthenosphere, leading to a depression of mantle

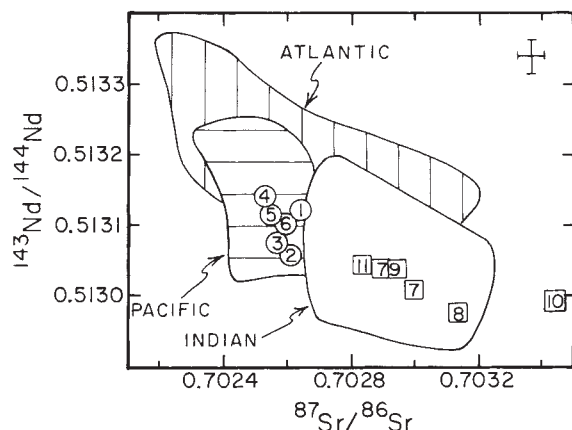


Fig. 4  $^{87}\text{Sr}/^{86}\text{Sr}$  versus  $^{143}\text{Nd}/^{144}\text{Nd}$ . Symbols and fields as in Fig. 3.

isotherms beneath the AAD. Thus, the AAD may represent a 'cold spot'.

Recent seismological observations are consistent with the presence of unusually cold mantle beneath the ridge axis in this region. Forsyth *et al.*<sup>25</sup> have reported that in the region south of Australia, Rayleigh wave phase velocities for crust <10 Myr are unusually fast beneath seafloor that shows a negative depth anomaly of more than 500 m, which includes a wide area centred on the AAD. The authors attribute these high surface wave velocities to anomalously high shear-wave velocities throughout the upper 200 km and particularly within the upper 30 km of the mantle. Similarly, global analyses of seismic wave travel-time anomalies have suggested uncommonly fast velocities at depths of 250–350 km within the mantle beneath the AAD (refs 26 and 27). These findings thus support inferences made on the basis of bathymetry that unusually cold mantle material is present beneath the region of anomalous depth centred on the AAD.

The major-element composition of these SEIR basalts also support the hypothesis that unusually cold mantle is present beneath the AAD (ref. 19; E.M.K., C.H.L. and H.S., manuscript in preparation). Klein and Langmuir<sup>20</sup> have suggested that regional variations in the major-element composition of basalts can be used to constrain variations in the extents and pressures of melting resulting from differences in the temperature of the sub-solidus mantle. During adiabatic upwelling, unusually cold mantle can be expected to intersect the solidus at a lower pressure and melt less upon ascent; oxides that are sensitive to the pressure of melting, such as FeO, will reflect this lower pressure of melting. Along-axis variations in FeO contents, normalized at a constant MgO content, for these SEIR basalts are shown in Fig. 2. Normalized FeO contents reach a minimum within the AAD, suggesting that these basalts are produced by lower pressures of melting, resulting from lower mantle temperatures beneath the AAD.

**Implications for the plan view of mantle convection beneath the SEIR.** The studies discussed above suggest that the AAD itself is anomalous compared with adjacent ridges, and our data suggest that a boundary between chemical provinces exists within the AAD. The question remains, then: how can a relatively sharp boundary (<200 km) between large-scale isotope provinces exist in a convecting mantle? One possible geometry might involve a 'direct convergence' pattern in which the boundary between two convective domains occurs beneath the eastern half of the AAD. In this scheme, an eastern convective domain continually feeds the Pacific Ocean isotope signature to the AAD from the east, while a western convective domain continually feeds Indian Ocean mantle to the AAD from the west. The isotope boundary would then coincide with the position

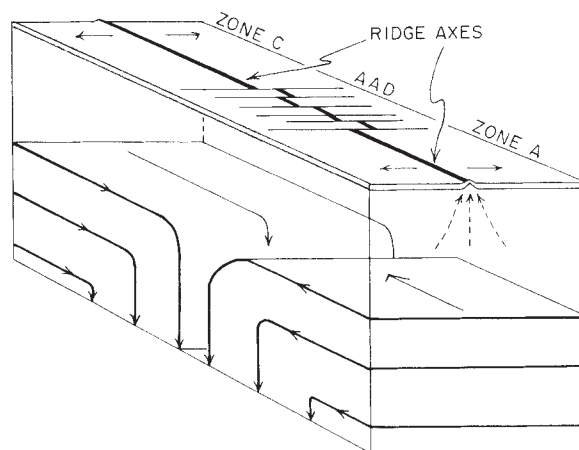


Fig. 5 Schematic model for large-scale, convective mantle downwelling beneath the AAD. Local upwelling associated with spreading of the ridge axis is indicated by the dashed arrows. See text for discussion.

along the ridge axis where the limbs of the two convective regimes converge and descend into the mantle. The descent of this material would lead to the postulated depression of mantle isotherms<sup>16</sup>, the anomalous depth within the AAD<sup>13</sup> and the apparent coldness of the mantle beneath it<sup>25</sup>. A schematic drawing of this model of the convection pattern beneath the SEIR is shown in Fig. 5.

Independent models of global patterns of mantle or melt flow have also predicted a convergence of east- and west-flowing sub-crustal material in the basin between Australia and Antarctica. Based on the assumption that continental roots restrict upper mantle flow to sub-oceanic paths, Alvarez<sup>28</sup> proposed that sub-Pacific mantle flows out through three gaps, one of which occurs between Australia and Antarctica. In addition, Vogt examined global topographical gradients along ocean ridges and suggested that sub-axial asthenosphere<sup>49</sup> or melt<sup>30</sup> converges beneath topographic lows, such as the AAD. Note also that an aeromagnetic study of this region<sup>17</sup> has identified the presence of two active and two extinct propagating rifts converging on the AAD from the east and west.

The model of mantle flow depicted in Fig. 5 calls upon a region of deep convective downwelling beneath a ridge axis where shallow upwelling, melting and accretion associated with seafloor spreading occur. Implicit in this model is a decoupling between the geometry of flow associated with deeper convection and that associated with upwelling and melting. Although the early observations of high heat flow and shallow bathymetry at ridges fostered the assumption that ridges are underlain by hot rising jets within the mantle, more recent analyses argue against this perspective (see, for example, refs 30 and 31) and suggest that seafloor spreading in at least some areas is a passive phenomenon that draws mantle material into the zone of melting. Thus, there need be little relationship between the plan view of deeper convection and the location of the ridges; a ridge can migrate over the surface of the Earth, in response to overall plate motions and interactions, and sample mantle of the composition and temperature beneath it at any given time, which is governed by deeper convection patterns.

There is some evidence, however, for a long-lived, close spatial association between the zone of downwelling and the ridge axis within the AAD. Reviewing the data on the bathymetry of the SEIR flanks, Weissel and Hayes<sup>16</sup> concluded that the processes that produce the observed depth anomalies along the ridge axis have persisted for at least 30 Myr, although the locus of



maximum depth has shifted a few degrees westwards with time. Furthermore, based on an analysis of sedimentary basin deposition in Australia and Antarctica, Veevers suggested that the processes responsible for the anomalous bathymetry in the region of the AAD have existed at least since the initial rifting of Australia and Antarctica during the Cretaceous<sup>32</sup> and possibly much longer<sup>33</sup>. The apparent longevity of this zone of downwelling and its association with the AAD is particularly remarkable in view of the fact that during the separation of Australia and Antarctica the ridge axis has migrated northwards more than 10° of latitude and probably eastward as well<sup>34,35</sup>. If, however, the zone of downwelling is a linear feature in plan view that trends roughly north-south, as we have depicted it schematically in Fig. 5, then the AAD would continually override this 'cold line' as the ridge axis migrates northward with time.

**The Indian Ocean isotope province.** Isotope variations in MORB have long been ascribed to mixing of ambient depleted upper mantle with one or more enriched, or less depleted, components, whose characteristics are known from the study of ocean-island basalts (see, for example, refs 36, 37). The peculiar isotopic character of both Indian Ocean MORB (Figs 3 and 4) and islands led Dupré and Allègre<sup>5</sup> and Hamelin *et al.*<sup>7</sup> to suggest that both the depleted and enriched components in the Indian Ocean are distinct in composition from those in the North Atlantic and Pacific. Hart<sup>38</sup> further suggested that, at least for the islands, one could define a globe-encircling band of anomalous compositions centred on 30° S latitude (the Dupal anomaly). The characteristics of this anomaly (high  $^{87}\text{Sr}/^{86}\text{Sr}$ ,  $^{207}\text{Pb}/^{204}\text{Pb}$  and  $^{208}\text{Pb}/^{204}\text{Pb}$  for a given value of  $^{206}\text{Pb}/^{204}\text{Pb}$ ) are reminiscent of an old continental source and could result from processes such as the subduction of sediment, metasomatism of the mantle wedge above a wet downgoing slab, or delamination of sub-continental lithosphere (see, for example, the discussion by Zindler and Hart<sup>37</sup>).

Regardless of the origin of the Dupal anomaly, it is important to note that no mixing trends have yet been observed in the Indian Ocean data set to suggest direct mixing of the hypothetical 'enriched' composition with a depleted end-member like that postulated for the North Atlantic and Pacific. This observation led Hamelin *et al.*<sup>7</sup> to suggest that thorough mixing in the past of the Dupal component with a depleted component like that evident in the North Atlantic and Pacific may have led to the development of a unique, depleted end-member in the upper mantle beneath the Indian Ocean. The fact that Indian Ocean MORB data do not show mixing trends emanating from a depleted composition like that seen in the North Atlantic suggests that on the scale sampled by melting, this depleted composition has been homogenized with the 'contaminant'. Thus, the apparently distinctive character of the depleted upper mantle beneath the Indian Ocean requires some mechanism that can both thoroughly contaminate the mantle beneath a large-scale region and yet maintain its convective isolation from the upper mantle beneath other ocean basins.

Our SEIR data define a geographical boundary to the Indian Ocean MORB province and, in so doing, point to an observation that may have bearing on the nature of the contamination event that produced the distinct Indian Ocean MORB signature. We note simply that all of the continents that presently circumscribe the Indian Ocean province were assembled and moved approximately as a unit, Gondwanaland, throughout the Palaeozoic. Thus, if upper mantle convection beneath these continental land masses was decoupled from the larger global convection patterns, and if contamination by delamination or subduction occurred during that time, then a mechanism for the production of a distinct Indian Ocean 'depleted' (but contaminated) MORB reservoir is possible. The sharp isotope boundary within the AAD is comprehensible from this perspective: in contrast to the separation of India and Africa from Antarctica, which in plan view opened an 'internal' ocean basin surrounded by Gondwanaland continents, the separation of Australia from Antarc-

tica opened a pathway from the Pacific province to the mantle beneath the Gondwanaland continents (see, for example, refs 34 and 35). Indeed, if convection within the Pacific Ocean mantle is bounded by continental roots<sup>28</sup> or subduction zones, then the region between Australia and Antarctica represents one of the few apparently unobstructed pathways for communication between Pacific and Indian Ocean (sub-Gondwanaland) convective regimes.

The existence of an eastern boundary to the Indian Ocean isotope province raises an obvious question concerning a possible western boundary. The location and characteristics of a western boundary, however, are not yet clear<sup>7</sup>. On Pb-Pb variation diagrams, some of the data of Hanan *et al.*<sup>39</sup> for the southern mid-Atlantic ridge significantly increase the degree of overlap between the fields for the Indian and Atlantic/Pacific Ocean MORB isotopic provinces. The distinctive compositions of these ridge axis samples, however, appear to be due to the influence of the nearby islands of Tristan da Cunha, Gough and the Discovery Tablemount which show the Dupal isotope signature. Pb isotope data are not available for the ridge segments east and west of the Bouvet triple junction in the South Atlantic, although samples from these ridges display Nd and Sr isotopic systematics that differ somewhat from those seen in the North Atlantic and, in addition, show trace element affinities with Indian Ocean MORB (refs 40 and 41). It is possible, therefore, that further sampling and analysis will extend the Indian Ocean MORB isotope province to the mantle beneath the South Atlantic. Indeed, this would support the observation made above regarding the possible insulation of sub-Gondwanaland mantle from external convective regimes, since South America and Africa, too, were united throughout the Palaeozoic. Like the breakup of Australia and Antarctica, the separation of Africa from South America opened a channel for communication with the sub-Gondwanaland upper mantle, and with time for convective mixing we might expect to see a mixture of the two MORB isotope signatures in the South Atlantic.

## Conclusions

Pb, Sr and Nd isotope compositions of basalts collected along the Southeast Indian Ridge (115° E–138° E) show considerable variations but, to first-order, form two distinct isotope groups that correspond to longitudinal position along the ridge axis. Samples from the western group show isotope characteristics commonly associated with Indian Ocean MORB whereas samples from the eastern group show isotopic affinities with North Atlantic/Pacific MORB. Based on these affinities and on proximity to either the Indian or Pacific Ocean ridge systems, we suggest that a fundamental boundary between different isotope provinces occurs beneath the Australian–Antarctic Discordance, a region of the seafloor which may overlie a zone of downwelling convective flow. We propose that the isotope boundary also results from the convergence of convective flow in which Indian Ocean-type mantle is continually fed to the AAD from the west and Pacific Ocean-type mantle is fed to the AAD from the east.

The authors thank D. Walker, D. Christie, J. Rubenstone, J. Mahoney, A. P. LeHuray, K.-H. Park, H. Brueckner, T. Plank, J. Weissel, R. Buck and N. Bogen for criticizing the manuscript. We also thank the scientific party and crew of R/V *Vema*, cruise 33, legs 1 and 2 for collecting samples. Supported by the National Science Foundation (C.H.L., A.Z., H.S., and L-DGO Deep Sea Sample Repository) and the Office of Naval Research (L-DGO DSSR). L-DGO Contribution number 4308.

Received 8 January; accepted 28 April 1988.

1. Subbarao, K. V. & Hedge, C. E. *Earth planet. Sci. Lett.* **18**, 223–228 (1973).
2. Hedge, C. E., Futa, K., Engel, C. G. & Fisher, R. L. *Contr. Miner. Petrol.* **68**, 373–376 (1979).
3. Sun, S.-S. *Phil. Trans. R. Soc. A297*, 409–445 (1980).
4. Cohen, R. S. & O'Nions, R. K. *J. Petrology* **23**, 299–324 (1982).
5. Dupré, B. & Allègre, C. J. *Nature* **303**, 142–146 (1983).
6. Hamelin, B. & Allègre, C. J. *Nature* **315**, 196–199 (1985).

7. Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **76**, 288–298 (1986).
8. Price, R. C., Kennedy, A. K., Riggs-Sneeringer, M. & Frey, F. A. *Earth planet. Sci. Lett.* **78**, 379–396 (1986).
9. Michard, A., Montigny, R. & Schlich, R. *Earth planet. Sci. Lett.* **78**, 104–114 (1986).
10. White, W. M., Hofmann, A. W. & Puchelt, H. *J. geophys. Res.* **92**, 4881–4893 (1987).
11. Ito, E., White, W. M. & Göpel, C. *Chem. Geol.* **62**, 157–176 (1987).
12. Mahoney, J. J. *et al.* *J. geophys. Res.* (submitted).
13. Hayes, D. E. & Conolly, J. R. in *Antarctic Oceanology II: the Australian–New Zealand Sector*, *Antarctic Res. Ser. vol. 19* (ed. Hayes, D. E.) 125–145 (Amer. Geophys. Union, Washington, D. C., 1972).
14. Weissel, J. K. & Hayes, D. E. *Nature* **213**, 518–521 (1971).
15. Weissel, J. K. & Hayes, D. E. in *Antarctic Oceanology II: the Australian–New Zealand Sector*, *Antarctic Res. Ser. vol. 19* (ed. Hayes, D. E.) 165–196 (Amer. Geophys. Union, Washington, DC, 1972).
16. Weissel, J. K. & Hayes, D. E. *J. geophys. Res.* **79**, 2579–2587 (1974).
17. Vogt, P. R., Cherkis, N. Z. & Morgan, G. A. in *Antarctic Earth Science* (eds Oliver, R. L., James, P. R. & Jago, J. B.) 608–613 (Australian Academy of Science, Canberra, 1983).
18. Anderson, R. N., Spariosu, D. J., Weissel, J. K. & Hayes, D. E. *J. geophys. Res.* **85**, 3883–3898 (1980).
19. Klein, E. M., Langmuir, C. H., Zindler, A. & Staudigel, H. *Trans. Am. geophys. Un.* **65**, 301 (1984).
20. Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **92**, 8089–8115 (1987).
21. Cohen, R. S., Evensen, N. M., Hamilton, P. J. & O’Nions, R. K. *Nature* **283**, 149–153 (1980).
22. Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175–195 (1984).
23. Manhès, G., Minster, J. F. & Allègre, C. J. *Earth planet. Sci. Lett.* **39**, 14–24 (1978).
24. Marsh, J. G., Brenner, A. C., Beckley, B. D. & Martin, T. V. *J. geophys. Res.* **91**, 3501–3506 (1986).
25. Forsyth, D. W., Ehrenbard, R. L. & Chapin, S. *Earth planet. Sci. Lett.* **84**, 471–478 (1987).
26. Dziewonski, A. M. & Anderson, D. L. *Am. Scient.* **72**, 483–494 (1984).
27. Woodhouse, J. H. & Dziewonski, A. M. *J. geophys. Res.* **89**, 5953–5986 (1984).
28. Alvarez, W. J. *geophys. Res.* **87**, 6697–6710 (1982).
29. Vogt, P. R. *Earth planet. Sci. Lett.* **29**, 309–325 (1976).
30. Houseman, G. A. *Earth planet. Sci. Lett.* **64**, 283–294 (1983).
31. McKenzie, D. & Bickle, M. J. *J. Petrology* (in the press).
32. Cande, S. C. & Mutter, J. C. *Earth planet. Sci. Lett.* **58**, 151–160 (1982).
33. Veevers, J. J. *Nature* **295**, 315–317 (1982).
34. Norton, I. O. & Sclater, J. G. *J. geophys. Res.* **84**, 6803–6830 (1979).
35. Morgan, W. J. in *The Sea Vol. 7* (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
36. Schilling, J.-G. *Nature* **242**, 565–571 (1973).
37. Zindler, A. & Hart, S. A. *Rev. Earth planet. Sci.* **14**, 493–571 (1986).
38. Hart, S. R. *Nature* **309**, 753–757 (1984).
39. Hanan, B. B., Kingsley, R. H. & Schilling, J.-G. *Nature* **322**, 137–144 (1986).
40. le Roex, A. P. *et al.* *J. Petrology* **24**, 267–318 (1983).
41. le Roex, A. P. *et al.* *Contr. Miner. Petrol.* **90**, 367–380 (1985).
42. Hamelin, B., Manhès, G., Albarede, F. & Allègre, C. J. *Geochim. cosmochim. Acta* **49**, 173–182 (1985).
43. Catanzaro, E. J., Murphy, T. J., Shields, W. R. & Garner, E. L. *J. Res. nat. Bur. Stand.* **72A**, (1968).
44. Macdougall, J. D. & Lugmair, G. W. *Nature* **313**, 209–211 (1985).
45. Tatsumoto, M., Knight, R. J. & Allègre, C. J. *Science* **180**, 1279–1283 (1973).
46. Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **67**, 340–350 (1984).
47. Macdougall, J. D. & Lugmair, G. W. *Earth planet. Sci. Lett.* **77**, 273–284 (1986).
48. Machado, N., Ludden, J. N., Brooks, C. & Thompson, G. *Nature* **295**, 226–228 (1982).
49. Vogt, R. R. & Johnson, G. L. *Earth planet. Sci. Lett.* **18**, 49–58 (1973).

# Primary structure and expression of a product from *cut*, a locus involved in specifying sensory organ identity in *Drosophila*

Karen Blochlinger, Rolf Bodmer, Joseph Jack\*, Lily Yeh Jan & Yuh Nung Jan

Howard Hughes Medical Institute and the Departments of Physiology and Biochemistry, University of California, San Francisco, California 94143, USA

\* Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

*In the absence of cut gene activity in Drosophila, external sensory organs are transformed into chordotonal organs. Here we show that the cut locus encodes a large protein containing a homoeodomain and is expressed in nuclei of cells in external sensory organs but not in cells within chordotonal organs.*

THE peripheral nervous system of *Drosophila* consists of a variety of sensory organs, each of which is composed of at least one neuron and the sensory structure formed by the support cells<sup>1–9</sup>. In the embryo, five distinct classes of peripheral sensory neurons can be identified according to their location, the sensory structures they innervate, the shape of their dendrites and their cell size<sup>2–5</sup>. Two of these classes consist of neurons with single dendrites<sup>4</sup>: the external sensory (es) neurons that innervate external sensory structures in the cuticle (for example, receptors for strain, touch and humidity); and the chordotonal (ch) neurons that innervate internal chordotonal structures (stretch receptors). The other three classes contain peripheral sensory neurons with multiple dendrites<sup>5</sup>.

The pattern of sensory organ distribution is invariant and provides a sensitive assay for the genetic analysis of the processes that specify sensory organ development. We recently observed that mutations of the *cut* locus affect the identity of es organs by causing them to differentiate as ectopic ch organs, without changing the overall number or position of sensory organs<sup>10</sup>.

We now present the primary structure of a large polypeptide from the *cut* locus deduced by cloning and sequencing DNA complementary to its messenger RNA. This polypeptide contains several distinctive features, including a domain of 61 amino acids with marked sequence similarity to the homoeodomain found in many products of developmental genes in *Drosophila* and other species<sup>11,12</sup>. We further demonstrate expression of the predicted *cut* protein in nuclei of cells within es organs but not in any cells of ch organs. These observations suggest that the region of the *cut* locus important for correct es organ differenti-

ation encodes a nuclear regulatory protein, possibly with DNA-binding properties similar to those attributed to other homoeodomain-containing proteins<sup>13,14</sup>.

## The *cut* locus

The *cut* locus is large and genetically complex<sup>15,16</sup> (Fig. 1). Mutations of the centromere-distal region are viable and affect the morphology of the wing (*cut wing*, *ct*) or the leg (*kinked femur*, *kf*), whereas mutations of the centromere-proximal region are embryonic lethal and cause the transformation of es organs into ch organs. There are two groups of embryonic lethal mutations, *lethal I* and *lethal II*. Mutations in the *lethal II* group are genetically null for *cut* function because they do not complement any mutations in the *cut* locus<sup>16</sup>, and they cause the same extent of es organ transformation as a large deficiency of the locus<sup>10</sup>.

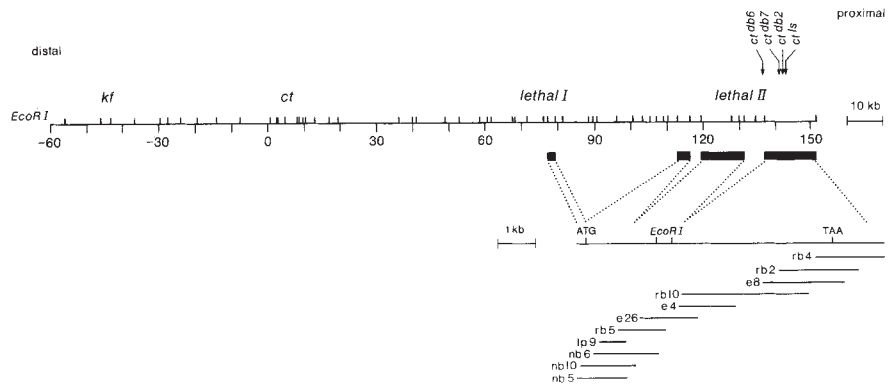
Over 200 kilobases (kb) of the *cut* locus have been previously cloned<sup>16</sup>. All previously known mutations that cause the es organ transformation map to the proximal half of the cloned region. To define the region necessary for es organ differentiation more precisely, we characterized 13 diepoxybutane-induced mutations which cause es organ transformation and are *lethal II* alleles. Three of these were found to have detectable genomic alterations within the *cut* locus. In each case the alteration is within the most proximal 20 kb of the cloned region (Fig. 1), suggesting that this region contains coding sequences of the *cut* gene product required for the normal development of external sensory organs.

We screened for and isolated DNA complementary to embry-



**Fig. 1** Genomic organization of the *cut* locus<sup>16</sup> including locations of diepoxybutane-induced mutations (*ct<sup>db2</sup>*, *ct<sup>db6</sup>*, *ct<sup>db7</sup>*), and *ct<sup>ls</sup>* (ref. 16), and cDNAs analysed (nb5, nb10, nb6, lp9, rb5, e26, e4, rb10, e8, rb2, rb4). Sites for *EcoRI* are indicated by notches. The two embryonic lethal regions *lethal I* and *lethal II* are marked. The black bars represent the *EcoRI* restriction fragments that hybridize to the isolated cDNAs. Intron-exon boundaries were not determined. Centromere-distal is left. Direction of transcription is from left to right.

**Methods.** All 13 diepoxybutane-induced mutations were *lethal II* alleles. Three mutations were associated with detectable genomic alterations within the *cut* locus as indicated in the figure. The position of these alterations was determined by genomic Southern mapping: *ct<sup>db7</sup>* is a small deletion of ~1 kb, *ct<sup>db2</sup>* and *ct<sup>db6</sup>* are probably also breakpoint mutations. Three embryonic cDNA libraries were screened for clones complementary to sequences in the *lethal II* region of the *cut* locus. Two of the libraries were oligo(dT)-primed<sup>43,44</sup>, the third was constructed in lambdaZAP (Stratagene) using both oligo(dT) and random priming of 8–21 h embryonic poly(A)<sup>+</sup> RNA. Screening for cDNA clones was performed according to published procedures<sup>45</sup>.



onic mRNA that hybridizes to this proximal region of the *cut* locus. From overlapping cDNAs, we determined the sequence of 8,217 base pairs (bp) of a *cut* transcript which spans more than 70 kb of genomic DNA from the proximal region (Fig. 1). Limited genomic sequencing confirmed that this transcript is derived from the *cut* locus. Fragments of this cDNA sequence hybridize to at least two transcripts of ~8.5 kb and ~9.5 kb in both early and late embryonic poly(A)<sup>+</sup> RNA (Fig. 2). Transcripts of these sizes are also present in poly(A)<sup>+</sup> RNA of pupae and adults (Fig. 2). Other mRNA species detected include a 5.3 kb transcript which is particularly prominent in pupae and an ~11 kb transcript found exclusively in adult heads (Fig. 2).

## Predicted protein sequence

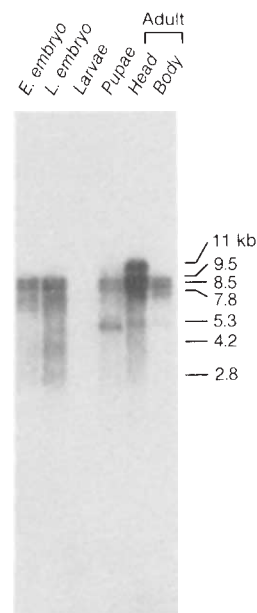
Deduction of the possible amino-acid sequences from the cDNA sequences predicts a long open reading frame bounded at both ends by non-coding sequences (Fig. 3). There is a consensus polyadenylation signal 11 bp upstream of the presumptive poly(A) tail. Of the two closely spaced AUG codons near the beginning of the open reading frame, the sequence immediately surrounding the first AUG conforms more closely to the consensus for ribosome binding sites in *Drosophila*<sup>17</sup> and may therefore be the translational start site. Assuming initiation of translation at the first AUG codon, the predicted primary translation product consists of 2,175 amino acids with an estimated relative molecular mass of 240,000 daltons. Some interesting features of the primary sequence are pointed out below.

A significant similarity to structures known as homoeodomains<sup>11,12</sup> is found near the carboxyl terminus of the predicted *cut* protein by comparative protein sequence analysis. Homoeodomains are stretches of 61 amino acids encoded by homoeoboxes which were first discovered in homoeotic genes of the *Antennapedia*-complex and the *bithorax*-complex in *Drosophila* and have since been found to be conserved in many other developmental genes in *Drosophila*, and in other species including vertebrates<sup>11,12</sup>. There is considerable circumstantial evidence that the homoeobox-containing genes encode regulatory proteins that bind to specific DNA sequences<sup>11–14</sup>. The *cut* homoeodomain appears to be the most divergent member of this family. Nevertheless, the nine amino acids which are invariant in all previously characterized homoeodomains<sup>11</sup> are unaltered in the *cut* homoeodomain (Fig. 4a).

Within the predicted *cut* protein, there are three internal repeats of 60 amino acids with 55–68% amino-acid identity (Fig. 4b). We name these *cut* repeats. The corresponding DNA sequences are also markedly similar (59–64% nucleotide identity), indicating that they may have arisen recently by duplication. We have not determined whether these amino-acid repeats

are encoded by different exons, like the repeats found in other proteins such as fibronectin<sup>18</sup>. The *cut* repeat appears to be unrelated to any previously determined protein structure.

There are four stretches of polyglutamate-aspartate in the predicted *cut* protein, ranging from 4 to 23 amino acids in length (Fig. 3, amino acids 271–293, 378–381, 547–554, 574–584, 2,071–2,077). Similar regions have been observed in other homoeodomain-containing proteins such as *engrailed*<sup>19</sup> and *deformed*<sup>20</sup>, and in the proteins encoded by the viral *myc* oncogene and its cellular homologues<sup>21</sup>; both *myc* and *engrailed* proteins have DNA-binding activity *in vitro*. The significance of



**Fig. 2** Transcription from the *lethal II* region of the *cut* locus. 5 µg of poly(A)<sup>+</sup> RNA from early embryos (1–14 h), late embryos (8–21 h), 3rd instar larvae, pupae, adult heads and adult bodies were electrophoresed, transferred to nylon membranes and hybridized to the most 5' *EcoRI* fragment of e26 (Fig. 1). All these mRNA species also hybridize to sequences at the 3' end of the cDNA sequence obtained (data not shown).

**Methods.** Poly(A)<sup>+</sup> RNA extraction, electrophoresis, transfer to nylon membranes, and hybridization conditions were previously described<sup>46</sup>. The probe was <sup>32</sup>P-labelled and was derived from specific priming of a single-stranded template. Transcript lengths were determined with a ladder of RNA standards (Bethesda Research Labs).

these exceedingly acidic sequences is unclear, but it has recently been shown that acidic regions are involved in activation of transcription<sup>22</sup>.

Many *Drosophila* gene products contain regions consisting almost exclusively of a single amino acid. These are predominantly encoded by identical trinucleotide repeats<sup>12</sup>. Similar repeats are also found in the *cut* cDNA sequence, resulting in regions rich in: glutamine and either histidine or alanine (Fig. 3, amino acids 194–211, 665–699, 1,147–1,161, 1,479–1,540); proline and alanine (amino acids 2,124–2,136); alanine (amino acids 194–202, 235–243, 616–630, 1,876–1,888, 2,004–2,014); and asparagine (amino acids 384–428). There is as yet no evidence for a functional significance of these regions in the *cut* protein or in other proteins.

## Distribution of predicted *cut* protein

Immunocytochemistry with whole-mount *Drosophila* embryos was performed using a polyclonal rat antiserum directed against a synthetic peptide that corresponds to an amino-terminal region of the predicted *cut* protein (Fig. 3). A repeated pattern of nuclear staining was observed below the peripheral epidermis of wild-type embryos (Fig. 5a). No nuclear staining was seen in the periphery of embryos deficient for the *cut* locus (Fig. 5e), suggesting that the staining is due to expression of the *cut* locus.

The arrangement of sensory organs is invariant<sup>2–5</sup> and most peripheral sensory neurons and their support cells can be distinguished according to their morphology and position using Nomarski optics (refs 4, 5 and Ghysen, A. & O'Kane, C. J., manuscript submitted) (Fig. 5b). Using these criteria, most of the labelled nuclei below the peripheral epidermis could be assigned to cells within es organs. As background for subsequent immunocytochemical experiments that sought to confirm this initial assignment, a brief description of the structure of es and ch organs follows.

A simple es organ contains a neuron and a set of support cells (typically three) that form concentric sheaths around the dendrite and produce the external sensory structure<sup>6–9</sup> (Fig. 5f). The innermost support cell (thecogen) forms a sheath around the tip of the dendrite, the middle support cell (trichogen) produces a sensory process (such as a sensory hair) and the outer support cell (tormogen) forms a circular socket or excavation at the base of the sensory process<sup>9</sup>. A chordotonal organ also contains a neuron and three support cells which form the elongated sheath<sup>6–9</sup>. It is presumed that the cells within a ch organ are homologous to the cells of an es organ<sup>7,10</sup>. Associated with the tip of the ch dendrite is a conspicuous refractile structure (scolopale) which corresponds to extensive arrays of microtubules and microfibrils within one of the ch support cells<sup>7,8</sup>. A similar but less extensive structure is seen within the innermost support cell near the tip of the es dendrite<sup>7–9</sup>. These probably correspond with the distinctive structures stained with the monoclonal antibody Mab21A6 described below.

The identity of the cells expressing *cut* was further investigated by labelling embryos simultaneously with an anti-*cut* serum and Mab21A6, a monoclonal antibody which stains only the scolopale of the ch organ and a dot near the tip of the es dendrite<sup>23,24</sup> and allows a clear distinction between es organs and ch organs<sup>10</sup>. Most of the peripheral cells labelled by the anti-*cut* antibodies are associated with the Mab21A6-stained dots of es organs. In no instance are labelled nuclei perceived in the cells forming scolopales or in any other cells associated with chordotonal organs (Fig. 5c, d).

To examine the distribution of anti-*cut* labelling within es organs, the anti-*cut* staining pattern was compared to the distribution of anti- $\beta$ -galactosidase staining cells in es organs in a line of transformant flies which appear to express  $\beta$ -galactosidase in all sensory organs (Chysen, A. & O'Kane, C. J., manuscript submitted). Upon double-labelling embryos from this line of flies with antibodies against  $\beta$ -galactosidase and with Mab21A6, four clearly marked cells are seen within a

simple es organ (Fig. 5g). Two of these cells lie immediately below the Mab21A6-stained dot and most likely correspond to the tormogen and trichogen cell. The neuron also expresses  $\beta$ -galactosidase and it can be identified by its dendrite, discernible with Nomarski optics, and by its position. Its identity can be further confirmed immunocytochemically; for example in embryos double-labelled with Mab21A6 and Mab21A4, a monoclonal antibody which stains the cytoplasm of all sensory neurons<sup>25</sup>, the dendrite of the es neuron can be easily traced to the Mab21A6-stained dot (Fig. 5h). The fourth cell expressing  $\beta$ -galactosidase is closely associated with the es dendrite and probably represents the thecogen cell. The same four cells of a simple es organ appear to be stained with the anti-*cut* serum (Fig. 5i, j). The nuclei of the trichogen and tormogen cells are always clearly labelled, whereas the staining in the neurons and the putative thecogen cells appears to be less intense and is often ambiguous. Further experiments with other antisera are necessary to adequately assess the extent of *cut* expression within individual cells of es organs.

All es organs in the embryo, including the antennamaxillary organ and the es organs in the spiracles, are labelled by the anti-*cut* antibodies. The earliest *cut* staining is seen before morphological differentiation of cells within peripheral sensory organs can be detected (stage 11–12, ref. 2), consistent with the contention that the *cut* locus controls the morphogenesis of es organs.

Besides es organs, other cells that are labelled with the anti-*cut* serum include some neurons with multiple dendritic arborizations<sup>5</sup> and cells lining malpighian tubules. This may be a reflection of the complexity of the *cut* locus, which affects several different tissues. It is also pertinent that a cDNA fragment containing the coding sequence of the peptide recognized by this anti-*cut* serum hybridizes to multiple species of embryonic, pupal and adult mRNA (Fig. 2). The pattern of *cut* expression at other developmental stages will be described elsewhere.

## Morphogenesis of es organs is directed by *cut*

Embryonic lethal mutations in the *cut* locus cause the morphological as well as antigenic transformation of es organs into ch organs<sup>10</sup>. Previous analyses with mosaic flies determined that *cut* activity is required either in or near es organs for them to acquire their correct identity<sup>10</sup>. In these studies, however, resolution at the cellular level was not possible. The observed immunocytochemical expression of *cut* in es organs (Fig. 5i, j) suggests that the *cut* gene function is autonomous to es organs.

The observation that the predicted *cut* protein is found in es organs but not in ch organs implies that at the time of onset of *cut* expression the cells are already determined to form es organs, that is, there must be other factors which influence the fate of es organs by regulating the expression of the *cut* locus. This expression precedes the detectable morphological differentiation of es organs, which further supports the hypothesis that the *cut* locus instructs cells to form es organs rather than ch organs, conceivably by inducing the expression of genes involved in the morphogenetic events of es organ differentiation.

The other *Drosophila* gene known to be involved in the specification of neuronal identity in the PNS that has been characterized molecularly is *sevenless*. Mutations in this gene cause the UV-sensitive photoreceptor cell, R7, to differentiate as a cone cell in each ommatidium of the imaginal eye disc<sup>26–28</sup>. The predicted *sevenless* protein has a strong structural similarity to several protein tyrosine kinases<sup>29</sup> and appears to be expressed in the cell membranes of several different cell types, including the photoreceptor cells and cone cells<sup>30–32</sup>. These results suggest that the *sevenless* protein may be a receptor for cell interactions in the imaginal eye disc. Unlike *sevenless* mutations, which cause a neuronal cell to be transformed into a non-neuronal cell, *cut* mutations cause transformations between two types of neurons along with their support cells<sup>10</sup>. Furthermore, the presence of a homoeodomain as well as the nuclear localization argues for a



[illegible]

**Fig. 3** Nucleotide sequence of *cut* cDNA and deduced amino-acid sequence of a *cut* product. Nucleotides and amino acids are numbered on the left. The homoeobox is boxed (amino acids 1,744–1,804). The *cut* repeats are underlined (amino acids 886–945, 1,339–1,398, 1,617–1,676). The sequence corresponding to the synthetic peptide used to generate the anti-*cut* serum is overlined (amino acids 567–581). There is a consensus polyadenylation signal AATAAA 11 bp upstream of the presumptive poly(A) tail.

**Methods.** Both strands of cDNA clones were sequenced in M13 derivatives by the dideoxy chain-termination method<sup>47</sup> using oligonucleotide primers and modified T7 polymerase (Sequenase, US Biochemicals). cDNA nb5 contains nucleotides (−268)–(+1,024); nb10, (−233)–(+1,270); nb6, 133–1,844; Ip9, 319–1,030; rb5, 809–2,086; e26, 1,387–1,839; e4, 2,440–3,912; rb10, 2,511–5,893; e8, 4,686–5,839; rb2, 5,090–7,262; rb4, 6,094–7,949. There is an extra T at position 6,704, and a T lacking at position 7,053 in rb4; e8 lacks the T at position 6,825. All the above deviations occur within non-coding sequences. The only difference observed within coding sequences is that the C at position 5,281 is missing in e8 (but present in rb10, rb2 and subcloned genomic DNA).

[illegible]

**Fig. 4** *a*, Alignment of the amino-acid sequences of the homoeodomains from *antennapedia* (*antp*)<sup>48</sup>, *engrailed* (*en*)<sup>43</sup>, *paired* (*prd*)<sup>49</sup> and *cut*. Amino acids which are identical among all these are boxed with solid lines. Amino acids which are identical between the *cut* sequence and at least one other are boxed with dashed lines. Asterisks denote the nine amino acids conserved in all homoeodomains characterized<sup>11</sup>. *b*, Alignment of the amino-acid sequences of the three *cut* repeats. Amino acids which are identical in at least two of the sequences are boxed.

regulatory function for the *cut* gene product, perhaps as a DNA-binding protein that regulates gene expression, as has been suggested for other homoeodomain-containing proteins<sup>11-14</sup>.

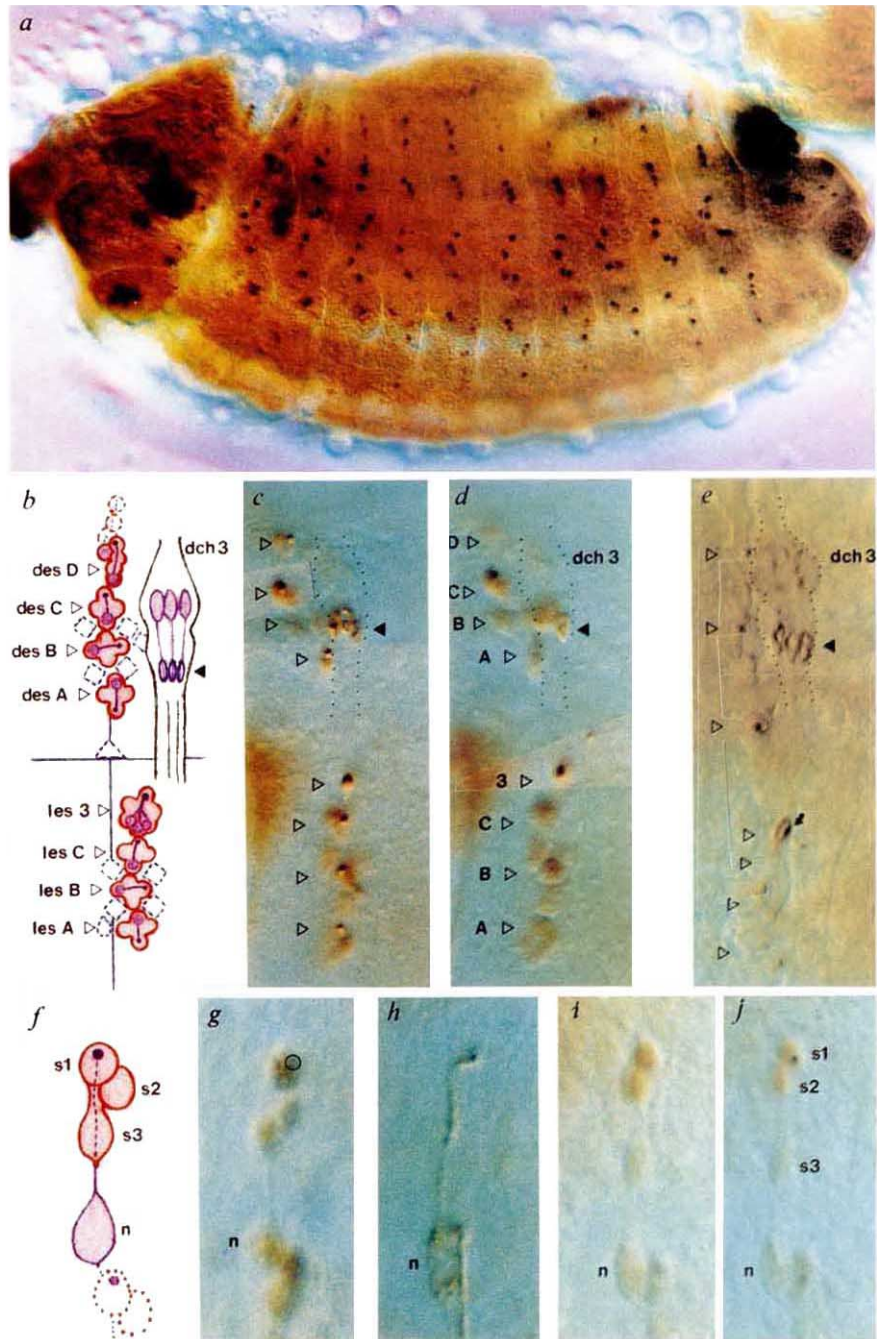
Each of the homoeobox-containing genes previously isolated in *Drosophila* is expressed in the nuclei of a unique subset of spatially restricted cells during early embryonic development<sup>11,12</sup>

and thereby specifies positional information. For example, transcripts and proteins of the segmentation gene *engrailed* have been localized in stripes with a periodicity of one segment at the cellular blastoderm stage<sup>19,33</sup>. It is speculated that different pathways of differentiation result from permutations of expression of homoeobox-containing genes in embryonic cells<sup>34,35</sup>. These patterns of gene expression probably involve



**Fig. 5** *a*, Staining pattern of a wild-type embryo with an antiserum directed against a synthetic peptide predicted from the *cut* cDNA sequence. Anterior is left and ventral is down. *b*, Diagrammatic representation of the identity and location of all the peripheral neurons in the dorsal and lateral cluster of a thoracic segment<sup>5</sup>. The es organs are shaded red, the ch organs are green, es neurons and ch neurons are shaded in purple, and the neurons with multiple dendrites are left blank. Nomenclature is according to Ghysen *et al.*<sup>4</sup>. In *b* through *d* blank arrowheads point to es organs and the black arrowhead points to the scolopales of ch organs. *c, d*, Two focal planes of the second thoracic segment of a wild-type embryo labelled with an anti-*cut* antibody and with Mab21A6. Using Nomarski optics, the outline of the cells containing anti-*cut* labelled nuclei in es organs could be traced in different focal planes to the Mab21A6-labelled dot. No staining was observed in the cells of ch organs, which were distinguished by their morphology and position and by the Mab21A6-stained scolopales. The position of the ch organs as seen by Nomarski optics is outlined with dots. The nuclei of the ch neurons and support cells, are dorsal to the scolopales. *e*, Staining pattern of the corresponding region in a homozygous mutant embryo deficient for the *cut* locus (*Df(1)ct<sup>4b1</sup>* refs 10, 15) that was labelled with an anti-*cut* antiserum and with Mab21A6. In the mutant embryo the es dots are transformed into scolopales as seen most clearly in the transformed *les3* organ (curved arrow). The transformation of the other es dots is not as evident in Fig. 5*e* because the scolopales are oriented perpendicular to this focal plane or are only partially transformed<sup>10</sup>. The positions of the transformed or partially transformed es organs are indicated with blank arrowheads, the position of the endogenous ch organs is outlined with dots, endogenous scolopales are marked by a black arrowhead. No nuclear staining is observed in the mutant embryo. *f*, Schematic representation of a simple es organ of the dorsal abdominal cluster (*desD*). The neuron (*n*) has a dendrite that extends to the Mab21A6-stained dot. There are probably three support cells in a simple es organ: the tormogen cell, the trichogen cell and the thecogen cell, which probably correspond to *s1*, *s2* and *s3*, respectively. *g, desD* of a 37a embryo labelled with anti- $\beta$ -galactosidase serum; the neuron is indicated (*n*). *h, desD* of an embryo stained with Mab21A4. *i, j*, Two focal planes of *desD* stained with anti-*cut*. Embryos *g, j*, were simultaneously labelled with Mab21A6 to indicate the position of the es dot. In (*g*) the Mab21A6 dot is out of focus and is marked by a circle.

**Methods.** The procedure for embryo collection, removal of the vitelline membrane, and staining of whole-mount embryos with antibodies is described elsewhere<sup>10</sup>. The concentrations of the antisera used were: anti-*cut* rat serum 1:1,000, Mab21A6 (obtained from J. Pollack and S. Benzer) 1:10, Mab21A4 1:10, anti- $\beta$ -galactosidase 1:500. The staining of the embryo shown in (*a*) was intensified<sup>50</sup>. 37a is a transformant line of flies provided by Drs C. O'Kane and W. Gehring. Integrated in the genome of these flies is a *lacZ* gene which is fused in-frame to the transposase gene of a P element<sup>51</sup> and which appears to be expressed in all cells of sensory organs (A. Ghysen, and C. J. O'Kane, manuscript submitted).



cross-regulatory interactions, possibly modulated by competition between different homoeodomain-containing proteins for similar *cis*-regulatory sequences<sup>13,14</sup>. In addition, many of the homoeobox-containing genes which specify positional information early in embryonic development have been found to be subsequently expressed in subsets of cells in the developing nervous system<sup>36</sup>. Indeed, it has recently been observed that the axon morphology of an identified CNS neuron is altered in the absence of expression of the homoeobox-containing segmentation gene *fushi tarazu*<sup>36</sup>. This later expression of homoeobox-containing genes appears to be controlled by different regulatory

interactions<sup>36</sup>.

The predicted *cut* protein also appears to be nuclear, but it differs from previously characterized homoeodomain-containing proteins in several aspects. The onset of *cut* gene expression as observed immunocytochemically is much later than that of the other cloned homoeobox-containing genes. Moreover, it is expressed in a 'cell-type' specific pattern rather than in a primarily 'position-specific' pattern. It also contains the most divergent homoeodomain yet identified and may therefore be part of a new family of homoeodomain-containing proteins.

It has recently been shown that a gene required for the differentiation of the touch receptor neurons in *C. elegans* also contains a homoeobox<sup>37</sup>. Moreover, several homoeobox-containing genes cloned in vertebrates seem to be primarily expressed in the nervous system<sup>38-42</sup>. These observations indicate that the relevance of homoeobox-containing genes in specifying cell fate in the nervous system may not be confined to *Drosophila*.

We thank Drs T. Kornberg and T. Hsieh for cDNA libraries,

Dr S. Benzer for Mab21A6, Drs C. O'Kane and W. Gehring for transformant line 37a, Drs A. Ghysen and C. O'Kane for results prior to publication, Dr T. Schwarz, S. Shepherd, Dr C. Turck and M. Doherty for technical assistance, and Drs W. Bender, A. Ghysen and P. O'Farrell for their comments. This study was supported by the Howard Hughes Medical Institute and the NIH. K.B. had a fellowship from the Fonds Nationaux Suisse, L.Y.J. and Y.N.J. are Howard Hughes investigators.

Received 5 April; accepted 22 April 1988.

1. Bate, C. M. in *Handbook of Sensory Physiology* Vol. IX (ed. Jacobson, M.) 1-53 (Springer, Berlin, Heidelberg, New York, 1978).
2. Campos-Ortega, J. A. & Hartenstein, V. in *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, Heidelberg, New York, 1985).
3. Dambly-Chaudiere, C. & Ghysen, A. *Wilhelm Roux Arch. dev. Biol.* **195**, 222-228 (1986).
4. Ghysen, A., Dambly-Chaudiere, C., Aceves, E., Jan, L. Y. & Jan, Y. N. *Wilhelm Roux Arch. dev. Biol.* **195**, 281-289 (1986).
5. Bodmer, R. & Jan, Y. N. *Wilhelm Roux Arch. dev. Biol.* **196**, 69-77 (1987).
6. Hertweck, H. *Z. wiss. Zool.* **139**, 559-663 (1931).
7. Zacharuk, R. Y. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 6 (eds Gilbert, L. I. & Kerkut, G. A.) 1-69 (Pergamon, New York, 1985).
8. McIver, S. B. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 6 (eds Gilbert, L. I. & Kerkut, G. A.) 71-132 (Pergamon, New York, 1985).
9. Hartenstein, V. *Development* (in the press).
10. Bodmer, R. *et al. Cell* **51**, 293-307 (1987).
11. Gehring, W. J. *Science* **236**, 1245-1252 (1987).
12. Scott, M. P. & Carroll, S. B. *Cell* **51**, 689-698 (1987).
13. Desplan, C., Theis, J. & O'Farrell, P. H. *Nature* **318**, 630-635 (1985).
14. Hoey, T. & Levine, M. *Nature* (in the press).
15. Johnson, T. K. & Judd, B. H. *Genetics* **92**, 485-502 (1979).
16. Jack, J. W. *Cell* **42**, 869-876 (1985).
17. Cavener, D. R. *Nucleic Acids Res.* **15**, 1353-1361 (1987).
18. Patel, R. S., Odermatt, E., Schwarzbauer, J. E. & Hynes, R. O. *EMBO J.* **6**, 2565-2572 (1987).
19. Kornberg, T., Siden, I., O'Farrell, P. H. & Simon, M. *Cell* **40**, 45-53 (1985).
20. Regulski, M., McGinnis, N., Chadwick, R. & McGinnis, W. *EMBO J.* **6**, 767-777 (1987).
21. Kassisi, J. A., Poole, S. J., Wright, D. K. & O'Farrell, P. H. *EMBO J.* **5**, 3583-3589 (1986).
22. Guarente, L. *Cell* **52**, 303-305 (1988).
23. Zipursky, S. L., Ventakesh, T. R., Teplow, D. B. & Benzer, S. *Cell* **36**, 15-26 (1984).
24. Pollack, J. A. *Calif. Inst. Tech. Biol. A. Rep.* **203**, 203 (1985).
25. Jan, Y. N., Ghysen, A., Christoph, I., Barbel, S. & Jan, L. Y. *J. Neurosci.* **5**, 2453-2464 (1985).

26. Harris, W. A., Stark, W. S. & Walker, J. A. *J. Physiol., Lond.* **256**, 415-439 (1976).
27. Campos-Ortega, J. A., Jurgens, G. & Hofbauer, A. *Wilhelm Roux Arch. EntwMech Org.* **186**, 27-50 (1979).
28. Tomlinson, A. & Ready, D. F. *Dev. Biol.* **123**, 264-275 (1987).
29. Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. *Science* **236**, 55-63 (1987).
30. Banerjee, U., Renfranz, P. J., Pollock, J. A. & Benzer, S. *Cell* **49**, 281-291 (1987).
31. Tomlinson, A., Bowtell, D. D. L., Hafen, E. & Rubin, G. M. *Cell* **51**, 143-150 (1987).
32. Banerjee, U., Renfranz, P. J., Hinton, D. R., Rabin, B. A. & Benzer, S. *Cell* **51**, 151-158 (1987).
33. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell* **43**, 59-69 (1985).
34. Garcia-Bellido, A. *Am. Zool.* **17**, 613-629 (1977).
35. Scott, M. P. & O'Farrell, P. H. *A. Rev. Cell Biol.* **2**, 49-80 (1986).
36. Doe, C. Q., Hiromi, Y., Gehring, W. J. & Goodman, C. S. *Science* **239**, 170-175 (1988).
37. Way, J. & Chalfie, M. *Cell* (in the press).
38. Awgulewitsch, A., Utset, M. F., Hart, C. P., McGinnis, W. & Ruddle, F. H. *Nature* **320**, 328-335 (1986).
39. Simeone, A. *et al. Nature* **320**, 763-765 (1986).
40. Carrasco, A. E. & Malacinski, G. M. *Dev. Biol.* **121**, 69-81 (1987).
41. Joyner, A. L. & Martin, G. R. *Genes Dev.* **1**, 29-38 (1987).
42. Odenwald, W. F. *Genes Dev.* **1**, 482-496 (1987).
43. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. *Cell* **40**, 37-43 (1985).
44. Nolan, J. M., Maxwell, P. L., Wyckoff, E. & Hsieh, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3664-3668 (1986).
45. Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. *Science* **237**, 749-757 (1987).
46. Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. *Nature* **331**, 137-142 (1988).
47. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
48. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. J. *EMBO J.* **5**, 733-739 (1986).
49. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033-1040 (1986).
50. Newman, G. R., Jasani, B. & Williams, E. D. *Microscopy* **132**, RP1-RP2 (1983).
51. O'Kane, C. J. & Gehring, W. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9123-9127 (1987).

# Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein

Ian A. Hope\*, Subramony Mahadevan & Kevin Struhl

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115, USA

*Derivatives of the yeast GCN4 transcription factor containing acidic regions of 35 to 40 amino acids fused directly to the DNA-binding domain are fully functional in vivo. High resolution deletion analysis and proteolytic mapping suggest that the activation region is a repeated structure composed of small units acting additively. Acidic character is a feature of the structural motif, possibly a dimer of  $\alpha$ -helices from two GCN4 monomers, that may be important for interactions with the basic transcriptional machinery.*

INITIATION and regulation of eukaryotic messenger RNA synthesis by RNA polymerase II requires transcriptional activator proteins that interact with specific promoter DNA sequences<sup>1-3</sup>. The DNA-binding functions of these activator proteins are localized to autonomous domains that contain less than 100 amino-acid residues<sup>4-8</sup>. Analysis of yeast GCN4 (ref 5) and GAL4 (refs 4, 9) proteins indicates that DNA-binding and transcriptional activation are independent functions carried out by distinct regions of the protein.

Surprisingly, the GCN4 and GAL4 transcriptional activation functions are performed by relatively short regions of acidic character that do not have rigid sequence requirements<sup>2,5,10</sup>.

Different portions of the GCN4 acidic region are equally capable of activating transcription, and there is no sequence homology between the GCN4 and GAL4 activation regions. Furthermore, acidity is the only obvious common characteristic of transcriptional activation regions selected from random segments of *Escherichia coli* DNA<sup>11</sup>, and mutations of a GAL4 derivative that increase or decrease activation usually increase or decrease negative charge, respectively<sup>12</sup>. In addition, the jun oncoprotein, a vertebrate transcription factor<sup>13,14</sup>, activates transcription in yeast through an acidic region<sup>15</sup>.

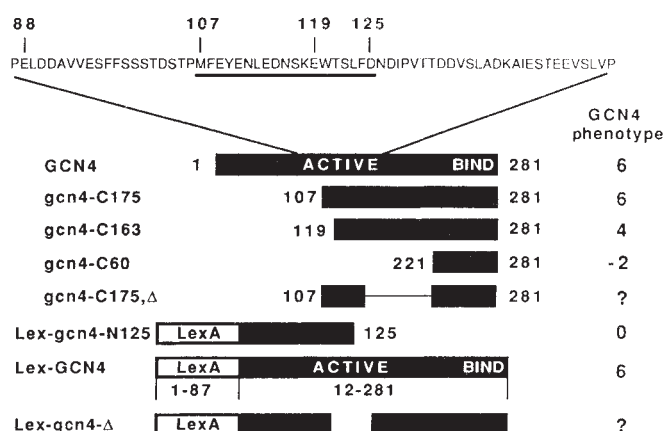
Although a growing body of evidence indicates the importance of acidic residues, little is known about the important structural characteristics of transcriptional-activation regions. This lack of understanding is due partly to the fact that no single activation region has been analysed in detail. Here, we investigate the

\* Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK



**Fig. 1** Structural and functional analysis of GCN4 and LexA-GCN4 derivatives. The structures of various GCN4 and LexA-GCN4 proteins are indicated by black bars with the N- and C-terminal residues defined as in the wild-type proteins. The wild-type GCN4 protein contains 281 amino acids and consists of a fully functional and dimeric DNA-binding domain (BIND) located within the 60 C-terminal amino acids<sup>5,20</sup>, and a centrally located transcriptional activation region (ACTIVE)<sup>4</sup> whose primary sequence (residues 87–152) is shown above. The 19 amino-acid segment implicated as being functionally important from the phenotypes of *gcn4-C175* and LexA-N125 is underlined. The GCN4 phenotypes are shown in an arbitrary scale, defined in the methods and used in the Tables, where 6 indicates wild-type activity, 4 indicates partial activity as observed for *gcn4-C163*, 0 indicates no GCN4, and –2 indicates the repression phenotype as observed for *gcn4-C60* (ref. 5).

**Methods.** The plasmids used to produce GCN4 and LexA-GCN4 proteins in yeast all derive from YCp88-GCN4 and YCp88-LexA-GCN4 (ref. 5), and contain the appropriate protein-coding region downstream of the constitutive *DED1* promoter in yeast/*E. coli* shuttle vector maintained at 1–2 copies per yeast cell. The GCN4 coding regions in YCp88-*gcn4-C175* and YCp88-LexA-GCN4 were modified to generate the internally deleted GCN4 derivatives whose structures and phenotypes are shown in Tables 1–3 and Fig. 2. For the experiment described in Table 1, YCp88-*gcn4-C175* DNA was cleaved with *KpnI* (corresponding to amino-acid residue 151), treated to various extents with *Bal31* nuclease and circularized. Upon introduction into yeast, derivatives conferring GCN4 function were selected by growth in the presence of 10 mM aminotriazole (see below) and subjected to DNA-sequence analysis. For the remainder of the experiments, appropriate DNAs were treated with *Bal31* nuclease, ligated to an 8 base pair *SalI* linker, and cloned into M13 derivatives. The deletion end-points were determined by DNA sequence analysis, and appropriate pairs of derivatives were combined to generate deletion mutants. The YCp88 DNAs were introduced into yeast strain KY803 (ref. 5), which is completely deleted for the GCN4 coding region, or KY330, an isogenic strain containing a plasmid which carries the *lacZ* gene downstream of a LexA binding site to permit measurement of transcriptional activation through the LexA DNA-binding domain. Levels of transcriptional activation through the GCN4 DNA-binding domain (total GCN4 activity) were determined by comparing growth rates on minimal plates supplemented with 20, 10 or 5 mM aminotriazole (AT) as described previously<sup>5</sup>. Growth in the presence of AT, which acts on the *HIS3* gene product to inhibit histidine biosynthesis, depends on active GCN4 to induce *HIS3* transcription, and the relative degree of AT resistance is directly related to the level of *HIS3* transcription<sup>25</sup>. To ensure reproducibility, all strains were tested on plates that were made at the same time, each derivative was tested on at least two independent occasions, several independent transformants for each derivative were tested, and cell growth was monitored by at least two individuals. The phenotypic classes are defined as follows: 6, growth at wild-type rates in 20 mM AT; 5, reduced growth in 20 mM AT, but wild-type growth in 10 mM AT; 4, slow growth in 10 mM, but wild-type growth in 5 mM AT; 3, slow growth in 5 mM AT, but normal growth in the absence of AT; 2 very slow growth in 5 mM AT, but normal growth in the absence of AT; 1 no growth in 5 mM AT, but nearly normal wild-type growth in the absence of AT. A phenotype of 0 is defined as identical to strains containing a GCN4 deletion mutation; such strains grow slowly in the absence of AT. GCN4 derivatives with a functional DNA-binding domain but lacking the transcriptional-activation regions repress transcription below the normal basal level; the growth of such cells is strongly inhibited on minimal plates without extra amino-acid supplements<sup>5</sup>. Phenotypes –1 and –2 represent mild or severe repression respectively. Transcriptional activation through the LexA DNA-binding domain was measured by standard plate and liquid culture assays of  $\beta$ -galactosidase activities.



structure and function of the GCN4 activation region by systematic and high resolution deletion analysis and by proteolysis of wild-type and deleted versions of the GCN4 protein.

## Minimal activation region

The initial experiment was designed to determine the minimal region sufficient for wild-type levels of transcriptional activation. Previously, *gcn4-C175*, the C-terminal 175 amino acids of GCN4, was shown to be the derivative with most deletions that retained full activity<sup>5</sup>. In the present experiment varying deletions within the *gcn4-C175* coding region were made by restriction cleavage between the sequences coding for the activation region and the DNA-binding domain, followed by treatment with *Bal31* nuclease. Derivatives still capable of activating transcription were identified by subjecting the collection of internal deletion mutants to a genetic selection (Fig. 1, Table 1).

The smallest fully functional derivative,  $\Delta 11$ , consists of 36 amino acids from the acidic region directly fused to a DNA-binding region containing 89 amino acids (Table 1). This indicates that a large, flexible spacer between the transcription activation region and the DNA-binding domain is not required. The minimal transcriptional activation region necessary for full function appears to be just 32 amino acids because the phenotype of a derivative containing only the N-terminal amino acids,  $\Delta 12$ , is indistinguishable from the wild-type protein, whereas all deletions that extend further towards the N-terminus of *gcn4-C175* show reduced transcriptional activation. Interestingly, the level of activation correlates very well with the length of the activation region but poorly with the overall size of the deletion or with the size of DNA-binding domain. The sole

exception,  $\Delta 16$ , may have less GCN4 activity because the deletion encroaches furthest into the DNA-binding domain. The shortest N-terminal segment with any detectable activity was the 25 amino-acid segment of  $\Delta 17$  and  $\Delta 18$ , suggesting that segments shorter than 25 amino acids are insufficient for transcriptional activation.

## Functional redundancy

Previous N- and C-terminal deletions indicated either that the 19 amino-acid segment between residues 107 and 125 is sufficient

**Table 1** Transcriptional activation by internal deletions of *gcn4-C175* selected for GCN4 function

| GCN4 derivative       | GCN4 amino acids present |            | Transcriptional activity |
|-----------------------|--------------------------|------------|--------------------------|
|                       | N-terminus               | C-terminus |                          |
| <i>gcn4-C175, Δ11</i> | 107–142                  | 193–281    | 6                        |
| <i>gcn4-C175, Δ12</i> | 107–138                  | 162–281    | 6                        |
| <i>gcn4-C175, Δ13</i> | 107–132                  | 164–281    | 5                        |
| <i>gcn4-C175, Δ14</i> | 107–134                  | 165–281    | 5                        |
| <i>gcn4-C175, Δ15</i> | 107–132                  | 176–281    | 3                        |
| <i>gcn4-C175, Δ16</i> | 107–137                  | 201–281    | 3                        |
| <i>gcn4-C175, Δ17</i> | 107–131                  | 200–281    | 2                        |
| <i>gcn4-C175, Δ18</i> | 107–131                  | 170–281    | 2                        |

The GCN4 amino acids which contribute to the N- and C-terminal segments are given for each derivative. The GCN4 activities are indicated on an arbitrary scale defined as in Fig. 1, where 6 indicates wild-type levels of activation, 0 indicates no activation, and –2 indicates repression.

|     | 116              | 123        | 130       | 137   | 144   | Amino acids | Net charge | GCN4 activity |
|-----|------------------|------------|-----------|-------|-------|-------------|------------|---------------|
|     | - -              | +-         | - -       | - -   | - +   | - -         |            |               |
| Δ35 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | EEVSL | wt          | -11        |               |
| Δ36 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | Egstk | 41          | -10        |               |
| Δ37 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | Egstk | 40          | -9         | 4             |
| Δ38 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | Edvq  | 40          | -11        |               |
| Δ39 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 38          | -7         |               |
| Δ40 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | Avdq  | 35          | -9         |               |
| Δ41 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 35          | -6         |               |
| Δ42 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 35          | -6         | 3             |
| Δ43 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 32          | -7         |               |
| Δ44 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 31          | -7         |               |
| Δ45 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 30          | -6         | 2             |
| Δ46 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 23          | -5         |               |
| Δ47 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 20          | -3         |               |
| Δ48 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 19          | -3         | 1             |
| Δ49 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 19          | -3         |               |
| Δ50 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 18          | -3         |               |
| Δ51 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 17          | -3         | 0             |
| Δ52 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 15          | -2         |               |
| Δ53 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 14          | -2         |               |
| Δ54 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 12          | -2         | -1            |
| Δ55 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 11          | -5         |               |
| Δ56 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 9           | -2         |               |
| Δ57 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 9           | -2         |               |
| Δ58 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 8           | -2         |               |
| Δ59 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grstk | 7           | -3         | -2            |
| Δ60 | MFEYENLEDNSKEWTS | SLFDNDIPVT | DDVSLADKA | IESTE | grpvk | 3           | +1         |               |

for the transcriptional activation or that GCN4 contains redundant activities<sup>5</sup>. As the data in Table 1 imply that the 19 amino-acid segment is insufficient for activation, it appears that residues towards either the N-terminus (as in LexA-gcn4-N125) or towards the C-terminus (as in gcn4-C175) contribute to a functional transcriptional-activation region. This redundancy was investigated by constructing internal deletions around and including the 19 amino-acid segment in LexA-GCN4 (Fig. 1), a bifunctional activator protein that can stimulate transcription through both the LexA and GCN4 DNA-binding domains<sup>5</sup>. Unlike the previous experiment in which functional derivatives were selected from a population, the derivatives analysed in Table 2 were generated by systematically fusing a series of N-terminal segments of LexA-GCN4 to a series of C-terminal segments. The resulting matrix of deletions were phenotypically analysed without being previously selected for GCN4 function.

The primary observation is that numerous deletions throughout the acidic region (amino-acid residues 88–147) still permit transcriptional activation by both DNA-binding domains. Many derivatives remove part or all of the 19 amino-

**Table 2** Transcriptional activation by both the LexA and GCN4 components of a series of internal deletions of LexA-GCN4

| GCN4 derivative | N-terminus |        | C-terminus | GCN4     | β-galactosidase |        |
|-----------------|------------|--------|------------|----------|-----------------|--------|
|                 | LexA       | GCN4   | GCN4       | Activity | Plate           | Liquid |
| LexA-gcn4-Δ19   | 1-87       | 12-14  | 85-281     | 6        | ++              | 350    |
| LexA-gcn4-Δ20   | 1-87       | 12-14  | 103-281    | 6        | ++              | 440    |
| LexA-gcn4-Δ21   | 1-87       | 12-92  | 103-281    | 6        | ++              | 450    |
| LexA-gcn4-Δ22   | 1-87       | 12-14  | 124-281    | -2       | -               | 6      |
| LexA-gcn4-Δ23   | 1-87       | 12-92  | 124-281    | 2        | +               | 85     |
| LexA-gcn4-Δ24   | 1-87       | 12-117 | 124-281    | 5        | ++              | 240    |
| LexA-gcn4-Δ25   | 1-87       | 12-14  | 133-281    | 0        | -               | 2      |
| LexA-gcn4-Δ26   | 1-87       | 12-92  | 133-281    | 0        | -               | 3      |
| LexA-gcn4-Δ27   | 1-87       | 12-117 | 133-281    | 0        | +               | 40     |
| LexA-gcn4-Δ28   | 1-87       | 12-129 | 133-281    | 4        | ++              | 170    |
| LexA-gcn4-Δ29   | 1-87       | 12-14  | 141-281    | -2       | -               | 8      |
| LexA-gcn4-Δ30   | 1-87       | 12-92  | 141-281    | 2        | +               | 30     |
| LexA-gcn4-Δ31   | 1-87       | 12-117 | 141-281    | 4        | ++              | 130    |
| LexA-gcn4-Δ32   | 1-87       | 12-129 | 141-281    | 5        | ++              | 400    |
| LexA-gcn4-Δ33   | 1-87       | 12-138 | 141-281    | 6        | ++              | 320    |
| LexA-gcn4-Δ34   | 1-87       | 12-95  | 144-281    | 2        | +               | 40     |
| LexA-GCN4       | 1-87       | 12-281 |            | 6        | ++              | 300    |

The amino-acid composition of each derivative is indicated in order from the N- and C-termini. The N- and C-terminal segments are joined by a *SalI* linker that encodes an additional arginine and proline residue. GCN4 activities are indicated in the arbitrary scale defined in Fig. 1 where 6 indicates wild-type levels of activation, 0 indicates no activation and -2 indicates repression. LexA activities were determined by measuring β-galactosidase activity produced from a *lacZ* gene downstream of a LexA binding site (see Fig. 1). For the β-galactosidase plate assay, hydrolysis of X-Gal gave either a strong blue (++), weak blue (+), or no blue (-) colour to yeast colonies. β-galactosidase activities for the liquid assay are defined as in previous work<sup>5</sup>.

**Fig. 2** High-resolution deletion analysis. For each derivative, the indicated amino-acid-sequence of the activation region was fused to a 100-residue GCN4 DNA-binding domain (at position 181). Residues shown in capital letters derive from the native GCN4 protein, whereas residues in small letters derive from the *SalI* linker and joint sequences. The number of GCN4 amino acids (excluding linker sequences), the net negative charge, and the level of GCN4 activity (determined as described in Fig. 1) are shown. Shown above the amino-acid sequences are the location of negatively and positively charged residues, and the probable boundaries that distinguish the phenotypic classes.

acid segment, yet still show full or partial transcriptional activation. Thus, even though the previous analysis attached critical importance to this region, analysis of internal deletions clearly reveals that this 19 residue segment is neither sufficient nor essential for transcriptional activation activity. Although these results could be complicated slight variations in the stability of derivatives *in vivo* (for which there are no data as GCN4 could not be detected biochemically), it is remarkable that so many derivatives are functional.

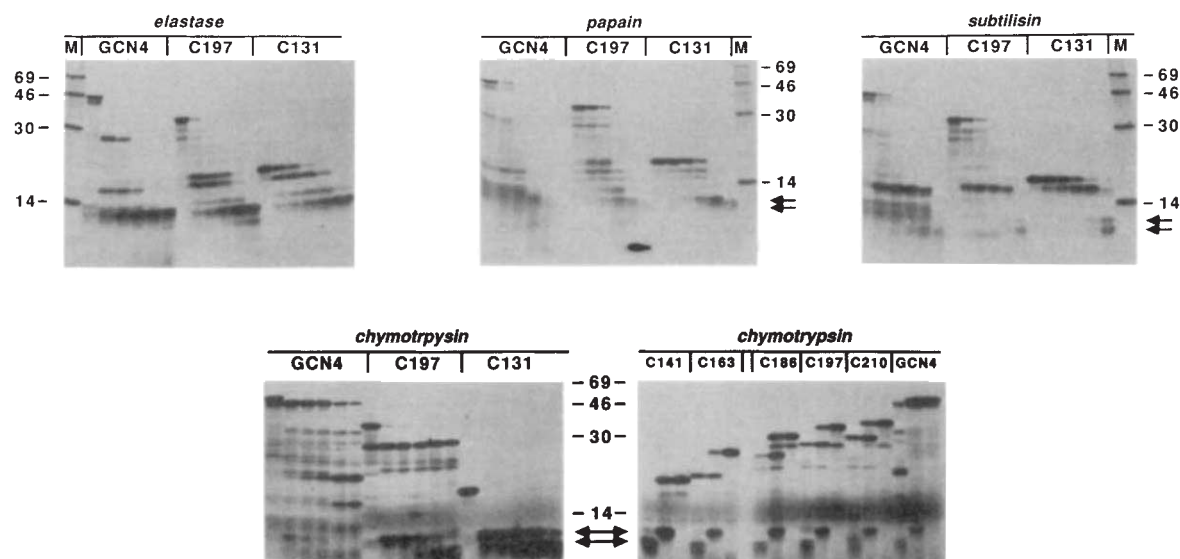
The level of transcriptional activation correlates fairly well with the amount of the acidic region that remains, but poorly to the particular regions that are retained. The exceptions are fusions to residue 133 (Δ25–Δ28), which have anomalously low activity. In addition, the region between residues 12 and 92 can impart some transcriptional activity in the absence of the primary acidic activation region (compare Δ23, Δ30, Δ34 with Δ22 and Δ29). This may be due to the short region between residues 50 and 70 that has the same proportion of acidic residues as the primary activation region.

## High resolution deletions

To investigate the nature of the transcriptional-activation region in more detail, we generated a large and sequential series of very fine scale internal deletions that could be easily related to each other. Constant C-terminal segments, containing the DNA-binding domain, were fused to a series of N-terminal segments ranging in size from 3 to 41 residues (Fig. 2). Although all the derivatives contain the 100 C-terminal amino acids, note that there are three classes of fusions representing the three possible reading frames; this affects the amino-acid sequence at the junction between the activation peptide and the DNA-binding domain. Thus, derivatives that contain the same number of GCN4 residues from the activation region can have a different number of acidic residues.

The phenotypes of the resulting derivatives fit a simple and striking pattern. Unlike structural elements such as active sites or domains in a protein, progressive deletion did not reveal a position where there was a sudden complete loss in activity. Rather, a series of small, step-wise reductions in activity was observed. Seven discrete phenotypes could be readily distinguished, ranging from high GCN4 activity for the largest derivatives to a transcriptional-repression phenotype when only the GCN4 DNA-binding domain is present. The range of activities observed and the lack of any derivatives which failed to fit into the simple pattern argue against any complications, such as effects on protein stability. Therefore, the different levels of activity observed almost certainly reflect differences in the amount of transcriptional-activation function retained by each derivative.





**Fig. 3** Protease digestion of GCN4 and deleted derivatives. For all the panels except for the bottom right (chymotrypsin), the indicated derivative was treated with 0, 1 $\times$ , 2 $\times$ , 4 $\times$ , 8 $\times$  or 16 $\times$  units of protease, where  $\times$  equals  $6.25 \times 10^{-4}$  for elastase,  $6.25 \times 10^{-5}$  for papain,  $10^{-5}$  for subtilisin, and  $10^{-4}$  for chymotrypsin. For the chymotrypsin experiment shown on the far right, each derivative was treated with 0, 1 $\times$  or 10 $\times$  units of chymotrypsin. The sizes of the cleavage products described below are approximate and do not always add up to the size of the initial substrates probably because of the very anomalous electrophoretic migration of GCN4 and its derivatives (GCN4 migrates at a position corresponding to a relative molecular mass ( $M_r$ ) of 45,000(45K) even though its true  $M_r$  is only 31K)<sup>16</sup>. Elastase initially cleaved GCN4 to a 16 K C-terminal segment present for all three derivatives, and an N-terminal segment of 26 K. The former was cleaved to a more stable 13.5 K segment. The C-terminal segment appeared less resistant to treatment with papain than to other proteases. C-terminal segments of greater stability than those derived from the N-termini could be discerned at 16 K, however, and then at 12.5 K. Note that N-terminal fragments would be expected to appear more intense than C-terminal fragments on the autoradiogram because of the distribution of the  $^{35}\text{S}$ -methionine residues (three are located at the extreme N-terminus at residues 1, 14 and 38, one is centrally located within the acidic region at 107, and one is located in the DNA-binding domain at residue 250). With subtilisin, the N-terminal segment did not appear as a single discrete fragment, whereas a stable C-terminal segment of 16 K was observed for all derivatives. For chymotrypsin, GCN4 was cleaved to 42 K N-terminal segment and a 12 K C-terminal segment which was cleaved further to 11 K. The arrows in figures correspond to bands representing the stable proteolytic fragments containing the GCN4 DNA-binding domain.

**Methods.** GCN4 and derivatives were generated by transcription and translation *in vitro* as previously described<sup>16</sup>. All protease digestions were of 0.5  $\mu\text{l}$   $\sim 10^4$  acid precipitable c.p.m. of  $^{35}\text{S}$ -methionine-labelled protein) of *in vitro* translation products in a final volume of 10  $\mu\text{l}$ . Reaction conditions were as follows: 10 mM Tris pH 8.8, 37  $^\circ\text{C}$  for elastase; 10 mM Tris pH 6.8, 25  $^\circ\text{C}$ ; for papain; 10 mM Tris pH 7.5, 37  $^\circ\text{C}$  for subtilisin; 10 mM Tris pH 7.9, 37  $^\circ\text{C}$  for chymotrypsin. After 45 min incubation, reactions were stopped by adding Laemmli sample buffer and heating to 100  $^\circ\text{C}$  for 5 min before applying to a 15 % SDS polyacrylamide gel for electrophoresis. The  $^{35}\text{S}$ -methionine-labelled protein fragments were revealed by autoradiography.

GCN4 activity appears directly related to the size of the transcriptional-activation region remaining. There is not a single case where a shorter region activates transcription more efficiently than a longer region. In contrast, there is no such precise relationship of transcriptional activity to the number of acidic residues. First, the boundaries that distinguish the phenotypic classes do not always correspond with the presence or absence of particular acidic residues. Second, there are several examples in which derivatives with fewer acidic residues activate transcription better than derivatives with more acidic residues. Third, there are several examples in which derivatives with an equal number of acidic residues fall into different phenotypic classes.

### Spacing requirements

Results presented here and elsewhere<sup>5,10</sup> suggest that the distance and orientation of the activation region with respect to the DNA-binding domain is functionally unimportant. To examine this more explicitly, several of the short N-terminal fragments described in the previous section were fused to a series of C-terminal segments of different sizes. The result was that for each N-terminal segment, fusion to C-terminal segments ranging from 61 to 130 amino acids produced similar phenotypes. This provides direct evidence that the activation region encodes an autonomous function that is equally active on a series of DNA-binding domains. In addition, it indicates that the distance between the activation region and DNA-binding domain is unimportant even at short distances, so that a spacer between

these two functions is not necessary (although the requirement for a very small spacer cannot be excluded).

One apparent anomaly is that all derivatives containing DNA-binding domains less than 110 amino acids show reduced levels of transcriptional activation when compared to the corresponding derivatives with DNA-binding domains of 130 residues. Of several possible explanations, perhaps the most probable is that the shorter DNA-binding domains are slightly defective for specific DNA-binding *in vivo*. In this regard, such derivatives have a slightly increased non-specific DNA-binding activity *in vitro*<sup>5</sup>. The wild-type phenotype of  $\Delta 11$ , which has an 89 residue DNA-binding domain, may represent an exception, especially because this derivative was pre-selected for wild-type function.

### Structural analysis

The fact that GCN4 can be extensively deleted without reducing its activity *in vivo* strongly suggests that tertiary structure is unimportant for the transcriptional-activation function. To investigate GCN4 structure in biochemical terms,  $^{35}\text{S}$ -labelled proteins<sup>16</sup> were treated with a series of proteases which preferentially cleave unstructured protein regions (Fig. 3). A protease-resistant C-terminal domain was readily released by treatment with elastase, subtilisin, or papain. Under the same conditions, these proteases completely cleaved the N-terminal portion of the protein, although with variation in pattern and rate. These results indicate that GCN4 contains an independently structured DNA-binding (C-terminal) domain with the remainder of the protein being relatively unstructured.

In contrast, cleavage by chymotrypsin generates two equally stable intermediates, suggesting the apparently contradictory view that the large N-terminal region is indeed structured. Similar results were obtained, however, with N-terminally deleted derivatives (gcn4-C210, C197, C186, C163); that is, large N-terminal segments of GCN4 were resistant to protease digestion as if they formed part of a structured domain, yet they could be removed without destroying the integrity of that domain. Interestingly, a stable N-terminal fragment was not generated for gcn4-C141 or -C131, as if deletion beyond residue 118 (the centre of the acidic activation region) leads to a sudden loss in the resistance of the N-terminal segment to chymotrypsin. Thus, the unusual pattern seen with chymotrypsin correlates completely with the presence of a functional transcriptional-activation region. One interpretation of this result is that the activation region has a local structure that inhibits chymotrypsin cleavage of the otherwise unstructured N-terminal region of GCN4.

## Molecular implications

Perhaps the most striking observation is that progressive deletion of the activation region causes a series of small, step-wise reductions of activity rather than defining a position where there is a precipitous loss of activity. This indicates that transcriptional-activation regions do not have a defined tertiary structure such as is found in active sites or domains in a protein, a view that is supported by our proteolysis experiments, by the short size and non-stringent amino-acid sequence requirements, and by the large number of protein deletions that remain functional. More importantly, the strong correlation between the length of the GCN4 activation region and the level of transcriptional activity is strongly suggestive of a repeating structure consisting of units which act additively.

Although acidic character is clearly involved in this activity,

it is not the only crucial feature because the level of transcriptional activation is only moderately correlated with net negative charge (Fig. 2). One clue to the structure of the GCN4 activation region is that the boundaries defining the step-wise levels of transcriptional activation appear to occur every seven amino-acid residues (residues 116, 123, 130, 137 and 144). The only exception to this rule is that derivatives in phenotypic classes 0 and 1 have endpoints within a single heptapeptide unit. Although this interpretation is clearly speculative, a repeat unit of seven amino acids is provocative because it could represent two turns of an  $\alpha$ -helix. The acidic region of GCN4 could form three  $\alpha$ -helices that are amphipathic, having acidic and hydrophobic residues along separate surfaces<sup>17,18</sup>.

Recently it was shown that a synthetic 15 amino-acid region whose sequence is consistent with forming two turns of an amphipathic helix could confer some transcriptional activity in yeast when fused to a 147 amino-acid GAL4 DNA-binding domain<sup>19</sup>. This result seems inconsistent with our observations that acidic regions of 15 residues do not possess transcriptional-activation function. The 147 residue GAL4 domain, however, contains 70 residues beyond those necessary for DNA binding, including an acidic region (net negative charge of -7 between residues 66 and 118). In our opinion, the 15 residue peptide probably does not represent an autonomous activation region, but possibly one or two of the basic units that can act in combination with the otherwise cryptic GAL4 acidic region. It should be mentioned that the major GAL4 activation region and several other yeast activation regions are unlikely to form amphipathic helices. Thus, although an amphipathic helical structure may be compatible with a functional transcriptional-activation region, a simple relationship between this structure and function appears unlikely.

From several considerations, we propose that the activation region is a dimer of intertwined  $\alpha$ -helices, one helix from each GCN4 monomer. First, the unusual chymotrypsin pattern reveals a structure in GCN4 that depends on the acidic activation region. Second, the stability of the dimeric DNA-binding domain, even in the absence of target DNA<sup>20</sup>, should facilitate the formation and/or stability of the proposed interactions between helices of different monomer subunits. Third, as the LexA DNA-binding domain binds very poorly to its operator because of weak dimerization<sup>21,22</sup>, it is probable that transcriptional stimulation through the LexA domain also requires that the activation region should facilitate dimerization<sup>5,11,23</sup>. The dimerization model also explains why amphipathic helices should form functional transcriptional-activation regions, as it would easily permit a structure involving interacting hydrophobic residues that are protected from solvent, and exposed acidic residues. Such amphipathic helices, however, would represent only one mechanism for forming dimers with exposed acidic residues; other mechanisms could be used for proteins such as GAL4.

As we and others have proposed, the activation region does not encode a catalytic activity but rather stimulates transcription by interacting with another protein. From various lines of evidence, we have argued against interactions with the basic histones and in favour of contacts with a TATA-binding protein or RNA polymerase II itself<sup>2,24</sup>. Presumably, such contact would cause an allosteric change in the general transcription machinery and hence would stimulate transcription. Although protein-protein interactions are usually thought to be highly specific, involving contact of complementary surfaces, the wide variety of sequences of transcriptional activation regions argues against an interaction of this type. Presumably, there is an ionic association between the acidic activation region and a basic pocket of the contacted protein. Perhaps the lack of specificity in this interaction is compensated for by the juxtaposition of the interacting elements along the DNA.

We thank Jennifer P. Macke for excellent technical assistance. This work was supported by the NIH.

**Table 3** Requirements for spacing between the transcriptional activation and DNA-binding domains for GCN4 activity

| GCN4 derivative        | GCN4 amino acids present |            | Transitional activity |
|------------------------|--------------------------|------------|-----------------------|
|                        | N-terminus               | C-terminus |                       |
| gcn4-C175, $\Delta$ 61 | 107-144                  | 148-281    | 6                     |
| gcn4-C175, $\Delta$ 62 | 107-144                  | 172-281    | 4                     |
| gcn4-C175, $\Delta$ 39 | 107-144                  | 179-281    | 4                     |
| gcn4-C175, $\Delta$ 63 | 107-144                  | 185-281    | 4                     |
| gcn4-C175, $\Delta$ 64 | 107-144                  | 201-281    | 4                     |
| gcn4-C175, $\Delta$ 65 | 107-144                  | 218-281    | 4                     |
| gcn4-C175, $\Delta$ 66 | 107-141                  | 148-281    | 6                     |
| gcn4-C175, $\Delta$ 41 | 107-141                  | 179-281    | 3                     |
| gcn4-C175, $\Delta$ 67 | 107-141                  | 201-281    | 3                     |
| gcn4-C175, $\Delta$ 68 | 107-136                  | 148-281    | 5                     |
| gcn4-C175, $\Delta$ 45 | 107-136                  | 179-281    | 2                     |
| gcn4-C175, $\Delta$ 69 | 107-136                  | 201-281    | 2                     |
| gcn4-C175, $\Delta$ 70 | 107-126                  | 148-281    | 3                     |
| gcn4-C175, $\Delta$ 47 | 107-126                  | 179-281    | 1                     |
| gcn4-C175, $\Delta$ 71 | 107-125                  | 148-281    | 3                     |
| gcn4-C175, $\Delta$ 49 | 107-125                  | 179-281    | 1                     |
| gcn4-C175, $\Delta$ 72 | 107-125                  | 201-281    | 1                     |
| gcn4-C175, $\Delta$ 73 | 107-121                  | 148-281    | 0                     |
| gcn4-C175, $\Delta$ 52 | 107-121                  | 179-281    | -1                    |
| gcn4-C175, $\Delta$ 74 | 107-121                  | 201-281    | -1                    |
| gcn4-C175, $\Delta$ 75 | 107-120                  | 148-281    | 0                     |
| gcn4-C175, $\Delta$ 53 | 107-120                  | 179-281    | -1                    |
| gcn4-C175, $\Delta$ 76 | 107-115                  | 148-281    | -1                    |
| gcn4-C175, $\Delta$ 56 | 107-115                  | 179-281    | -2                    |
| gcn4-C175, $\Delta$ 77 | 107-109                  | 148-281    | -2                    |
| gcn4-C175, $\Delta$ 60 | 107-109                  | 179-281    | -2                    |

The structure and GCN4 activity of each derivative is described as in Fig. 1, where 6 indicates wild-type levels of transcriptional-activation, 0 indicates no activation, and -2 indicates repression. The N- and C-terminal segments are joined by a *Sal*I linker that encodes an additional arginine and proline residue.



Received 14 March; accepted 25 April 1988.

1. McKnight, S. L. & Tjian, R. *Cell* **46**, 795-805 (1986).
2. Struhl, K. *Cell* **49**, 295-297 (1987).
3. Guarente, L. A. *Rev. Genet.* **21**, 425-452 (1987).
4. Keegan, L., Gill, G. & Ptashne, M. *Science* **231**, 699-704 (1986).
5. Hope, I. A. & Struhl, K. *Cell* **46**, 885-894 (1986).
6. Hollenberg, S. M., Giguere, V., Segui, P. & Evans, R. M. *Cell* **49**, 39-46 (1987).
7. Miesfeld, R., Godowski, P., Maler, B. & Yamamoto, K. *Science* **236**, 423-427 (1987).
8. Kumar, V. *et al.* *Cell* **51**, 941-951 (1987).
9. Brent, R. & Ptashne, M. *Cell* **43**, 729-736 (1985).
10. Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
11. Ma, J. & Ptashne, M. *Cell* **51**, 113-119 (1987).
12. Gill, G. & Ptashne, M. *Cell* **51**, 121-126 (1987).
13. Bohmann, D. *et al.* *Science* **238**, 1386-1392 (1987).
14. Angel, P. *et al.* *Nature* **332**, 166-171 (1988).
15. Struhl, K. *Nature* **332**, 649-650 (1988).
16. Hope, I. A. & Struhl, K. *Cell* **43**, 177-188 (1985).
17. vonHeijne, G. *EMBO J.* **5**, 1335-1342 (1986).
18. Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H. & Schatz, G. *EMBO J.* **5**, 1327-1334 (1986).
19. Giniger, E. & Ptashne, M. *Nature* **330**, 670-672 (1987).
20. Hope, I. A. & Struhl, K. *EMBO J.* **6**, 2781-2784 (1987).
21. Little, J. W. & Hill, S. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2301-2305 (1985).
22. Hurstel, S., Granger-Schnarr, M. & Schnarr, M. *EMBO J.* **7**, 269-275 (1988).
23. Struhl, K. *Cell* **50**, 841-846 (1987).
24. Chen, W., Tabor, S. & Struhl, K. *Cell* **50**, 1047-1055 (1987).
25. Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234**, 451-457 (1986).

## LETTERS TO NATURE

### The radio rings of Hercules A

Andrea Mason\*, Philip Morrison† & Alberto C. Sadun‡

\*Department of Physics, University of Padova, Italy

† Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ Bradley Observatory, Agnes Scott College, Decatur, Georgia 30030, USA

Hercules A is a powerful, double radio source, about 0.5 Mpc across. Although its two lobes differ, they share a family resemblance in overall form. Along the axis of the somewhat weaker west lobe, however, a unique structure is found: a regular sequence of broken circular rings of enhanced radio intensity. The five rings fit a simple kinematic model in which the rings are projected spherical shells of higher radio brightness. They grew as spherical waves with a uniform speed of expansion, all from one common place of origin, roughly where the narrow input jet widens sharply. The shells drift one after another outward at constant speed with the radiating medium of the lobe. Here we propose a model of inelastic proton-proton collisions within dense filamentary condensations of the input jet for the five initiating particle bursts. Source development contrasts sharply with that inferred for the prototype radio double, Cygnus A, where the radio lobe is nearly stationary and the electrons are newly energized at the outermost end of the lobe. The two sources point to quite distinct mechanisms for generating large radio-emitting plasmas.

The extended double radio source Hercules A(3C348) centres on a galaxy whose nucleus is active both in radio and in optical light. Two long narrow collinear radio-emission features (we call them jets), one on each side of the centre, connect the galaxy to much wider and longer radio-bright regions. These roughly continue along the jet axis, although with some irregularity of form and moderate curvature; we call them radio lobes. The source has been imaged at 5 GHz with the Very Large Array (VLA), its intensity pattern represented both by contour and by grey-scale maps of high spatial resolution over three decades of dynamic range<sup>1</sup>. This superb mapping has disclosed that along the west lobe there is an array of five incomplete, thin, remarkably circular rings of enhanced radio brightness. The farther out the ring, the larger is its diameter. (The smallest arc is less closely circular; it is partially filled like the next larger ring.) We have fitted the nearly round intensity contours roughly to five circular annuli, each having a similar width but a distinct centre and radius. We take these bright rings as real enhancements of radio intensity. The most distant point of the final arc is ~300 kpc from the central galaxy, as projected. The rings do not seem to be artefacts of aperture synthesis; scattering and related illusory images are physically implausible for radio features.

A simple kinematic interpretation is evident. One after the

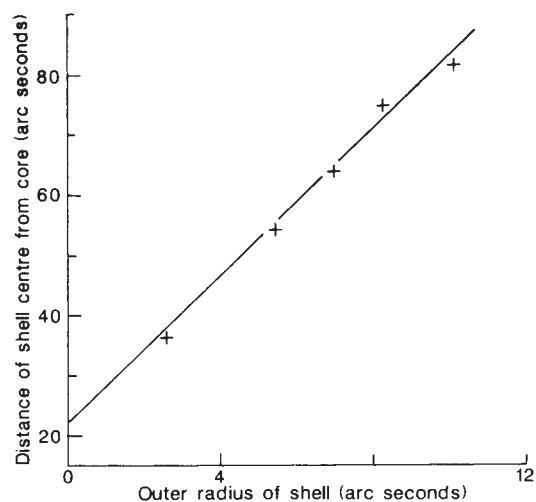


Fig. 1 The shell centre position and shell diameters for all five radio shells of Her A. Their position of common origin is about 20-25 arcsec from the core; the radiating medium of the lobe drifts uniformly outward at 6 times the common shell expansion speed. Comparison with other sources suggests that the drift speed is of the order 0.1 or 0.2 c h<sup>-1</sup>. We have used the values  $H_0 = 50$  h km per s-Mpc, with  $q_0 = 0.5$ . The observed redshift for Her A is  $z = 0.15$ , so that on the scale assumed 1 arcsec = 3.3 per h kpc projected on the sky.

other, each of the shells in turn began to expand, all originating close to a single position, near the plane where the narrow jet first abruptly widens into the lobe. Then each ring grew uniformly in diameter while its centre drifted steadily outwards along the lobe. Figure 1 records the visual fit of ring sizes and positions to the model. No counterpart rings are seen in the brighter east lobe (3/4 of the total radio power), although suggestive disturbances in its form are found at distances from the centre that roughly correspond to the ring distances out along the west lobe<sup>1</sup>.

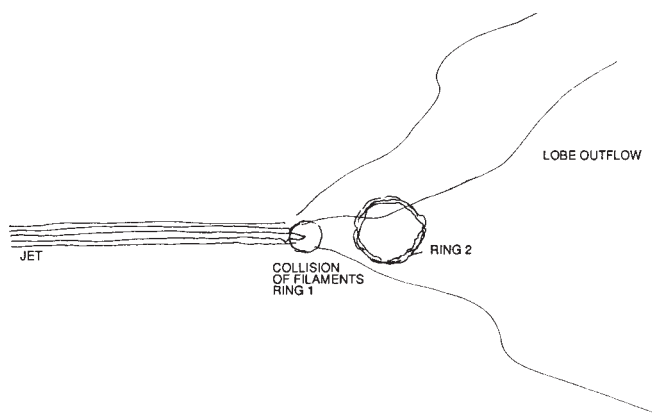
The implied history of the overall form of Hercules A is as follows. The tiny central engine sends out its bipolar jets of mildly relativistic protons with magnetic field. The west jet includes some fine-grained threads with current density far above average. At those long-stationary places where the lobes begin, the beam collimation was from time to time lost, perhaps by the cumulative effects of entrainment of ambient mass during the jet passage from the centre. The resemblance between east and west lobe morphology speaks against an entirely external dense target encountered by chance in the west jet; both beams vary internally with time in some centrally determined way, perhaps by transient increases in jet current density. Both narrow beams grow unstable as they propagate outward, lose their

collimation and quickly widen to lobe form. The wide lobes then lengthen steadily over time, as each drifting plasma lobe is fed both with energy and axial momentum by its own jet. Little energy need be lost in the jet slowing-down process, as long as the entrained mass that shares the axial momentum is large compared to that flowing in the incoming beam<sup>2</sup>. Then along the west lobe five shells formed one after another, each initiated by a burst of relativistic particles made within a small space-time volume by a collision between two dense proton filaments of the jet. The shells, simply spherical wavetrains in a uniform medium, expand uniformly while they are carried outward as part of the drifting medium.

We postulate a powerful burst of relativistic particles near where the jet first widens to initiate each shell. The burst arises when two intense jet filaments meet at modest collision angle. Each filament is a dense beam of mildly relativistic protons. Their nearly parallel collimation had been disturbed in the process of jet widening. Provided the space-time volume for the entire nuclear collision process is sufficiently small, something close to the radio shell thickness seen, an effectively point-like source of charged nuclear collision products will swiftly appear. These are secondary electrons and positrons with  $E/mc^2 \approx 10^3$ , the usual energy for synchrotron radio-emitting electrons. That expanding cloud of relativistic particles transfers a portion of its kinetic energy to the ambient field-threaded plasma, in which it excites an expanding spherical wave train of incremental density contrast (and perhaps also of magnetic field) within the lobe plasma. Each expanding shell appears projected as a ring of increased radio brightness that grows from its centre at the local wave speed. Meanwhile that centre is swiftly convected outward, as it shares the uniform axial motion of the lobe medium as a whole. (These plasma waves are expected to begin as weak shocks related to longitudinal magneto-acoustic modes, with speed rather like the transverse Alfvén waves.) Excess exciting particles left at the place of shell origin soon diffuse away, to merge into the lobe radio background. Signs of central emission are still seen within the two smallest shells, and the very end of the narrow jet bears a bright spot where possibly a new shell is incipient.

Entrainment of a large mass of ambient gas allows the jet to lose little bulk energy while giving up essentially all its forward momentum to the outwardly drifting lobe plasma. The narrow jet must enter the lobe region with proton energy  $E/mc^2 \approx 2-5$  to provide adequate production of electron secondaries; in the simplest case this will also be its bulk motion. The major portion of the jet might be decollimated to transfer its bulk kinetic energy to the radiating electrons of the entire lobe in a number of ways, including collisionless magnetic shock processes entirely apart from the single-particle nuclear collisions modelled to initiate the rings. Only a small ring-generating fraction need share the quasi-explosive localization of nucleon collisions. A generous estimate of the enhanced energy of the shells implies that the total energy carried in by the jet is larger than the wave energy in the shells by two orders of magnitude, which allows for inefficiencies in the energy transfer from beam protons to shocks in the lobe medium. The shell events do imply very high local proton current densities along intricate particle paths, presumably magnetically curled up within a few kiloparsecs. Sufficient energy for the events is expected in such jets; the surprise is that there is enough target matter to yield nuclear collisions at plausible efficiencies, up to  $\sim 10 \text{ g cm}^{-2}$  being required. This is perhaps explicable by such beam self-collision, provided the unseen fine colliding filaments are denser than the mapped image of the jet by some  $10^6$ , implying a local filament diameter on the parsec scale. That is still much larger than the expected scale of the jets at their origin deep within the active galaxy.

Thus Hercules A grew at the ends of its two narrow jets by steady axial extension of the wide lobes. No rings are seen on the east, though weak ring forms might be concealed within the brighter lobe contours. (Transit time delay might reduce the



**Fig. 2** A schematic drawing (not to scale) of the proposed collisional origin of the radio shells (see text). The narrow axial jet enters from the left. It is made up of many unresolved dense proton filaments. The jet continues until some instability causes it to lose collimation and widen into the much more uniform lobe. Near that place two filaments by chance collide. The energy and pressure of the charged collision secondaries generate in the lobe medium a wave front (ring 1) that expands isotropically while it is being convected downstream along with the bulk lobe flow. An earlier ring 2 has already expanded a good deal as it drifted downstream.

visibility of any fading eastern rings.) Our narrative implies intense transient gamma-ray emission (along with neutrinos) from such radio sources. A mean power of  $10^{45} \text{ ergs s}^{-1}$  for  $10^4$  or  $10^5$  years is needed to form the rings, the energy equivalent of a supernova explosion every few years. Thin-source thermal X-rays are currently seen over a large region about the source, perhaps external to it. Lobe speeds of comparable sources suggest plasma temperature and density in the range that might contribute most of the X-ray emission, but the medium is a complex one, and the plasma conditions are not easy to estimate.

One other large double source, Cyg A, has also been imaged in revealing detail<sup>3</sup>. The contrast between the inferred development of the two sources is striking, even though the sources have similar overall form and equipartition energy. (The more distant source, Hercules A, has a lobe volume larger by an order of magnitude.) These sources have long been classified with two distinct types of extended doubles; Cyg A is strongly edge-brightened, Her A somewhat edge-dimmed. The radio fine structure in Cyg A, many hair-like filaments smoothly folded into the complex lobes, suggests lobe structure there at a scale not usually resolved.

In Cyg A, the observed flattening of the radio emission spectrum with increasing distance out along each of the lobes from the centre has been used<sup>4</sup> to estimate local radiative ages for the lobe electrons. The most recently accelerated electrons lie at the outermost hotspots, at the terminus of the lobe, about where the jet in Cyg A presently encounters undisturbed ambient material. In Cyg A, the radiating lobe plasma seems nearly stationary with respect to the central galaxy; only the axial jet steadily moves outward against its contact surface with the ambient medium, generating nearly stationary plasma, some of which may even flow slowly backwards as an external 'cocoon'. The axial jet steadily extends, widens, and even undulates gently. A dozen or more edge-brightened doubles show<sup>5,6</sup> the same spectral regularity. Their spectra flatten as the terminal radio hot spots are approached along the lobe; innermost radio electrons have the steepest energy spectrum, and so appear the oldest.

In Her A the drifting rings indicate a development in time opposite to that of Cyg A. There are two major differences. The convected rings originate at and grow outwards from the innermost portion of the lobe, where it first widens. The most recently energized electrons should be there, not at the far end,



where the rings and lobe material have moved farthest and are oldest. (No map of the spectral index variation in Her A is known to us for a direct test). Again, Cyg A shows no sign of a large volume of outwardly moving radio lobe plasma. Only its distinct and axial jets can be said to move outwards. But in Her A, the rings serve to mark the steady outward axial drift of the overall medium of the lobe. No single model of jet acceleration processes is likely to account for two such distinct histories.

The bright radio source 3C310 also shows large ring-like radio structures. Its circular arcs appear to be built up of multiple overlapping segments, quite unlike the distinct sequential shells of Herc A. They require separate study.

We thank John Dreher of MIT for his maps and Dana Roberts of NASA-Goddard Space Flight Center for discussion.

Received 9 December 1987; accepted 12 April 1988.

1. Dreher, J. W. & Feigelson, E. D. *Nature* **308**, 43–45 (1984).
2. Morrison, P., Roberts, D. & Sadun, A. *Astrophys. J.* **280**, 483–490 (1984).
3. Perley, R. A., Dreher, J. W. & Cowan, J. J. *Astrophys. J. Lett.* **285**, L35 (1984).
4. Spinks, M. J., Rees, W. G. & Duffett-Smith, P. J. *Nature* **319**, 471–473 (1986).
5. Myers, S. T. & Spangler, S. R. *Astrophys. J.* **291**, 52–59 (1985).
6. Högbom, J. A. *Astr. Astrophys.* **36**, 173–192 (1979).

## Properties of high-density binary mixtures and the age of the Universe from white dwarf stars

E. Garcia-Berro\*†‡, M. Hernanz§†‡, J. Isern|| & R. Mochkovitch¶

\* Departament de Física Aplicada (Universitat Politècnica de Catalunya), C/Jordi Girona Salgado, 08028 Barcelona, Spain  
† Grup d'Astrofísica (Institut d'Estudis Catalans), Diagonal 647, 08028 Barcelona, Spain

‡ Departament de Física de l'Atmosfera, Astronomia i Astrofísica (Universitat de Barcelona), Diagonal 647, 08028 Barcelona, Spain

§ Departament de Física i Enginyeria Nuclear (Universitat Politècnica de Catalunya), Diagonal 647, 08028 Barcelona, Spain  
|| Instituto de Astrofísica de Andalucía (C.S.I.C.), Apdo 2144, 18080 Granada, Spain

¶ Institut d'Astrophysique de Paris, 98 bis Bd Arago, 75014 Paris, France

The luminosity of white dwarf stars can be attributed to the cooling process of their degenerate cores. The simple relationship existing between their luminosity and their age, together with the lack of white dwarfs fainter than  $\log(L/L_\odot) \approx -4.5$ , provides a method of measuring the age of the disk and consequently that of the Universe<sup>1,2</sup>. Winget *et al.*<sup>2</sup> have derived an age of the galactic disk of 9.3 Gyr and have then found that the Universe is young: 10.3 Gyr. These values depend on the assumption that completely ionized carbon and oxygen (the most abundant elements in white dwarf interiors) are miscible in solid phase. It is possible, however, that completely ionized carbon and oxygen separate during the process of crystallization<sup>3,4</sup>. The consequences of this behaviour on the evolution of mass-accreting carbon–oxygen white dwarfs in stellar binary systems<sup>5–7</sup> and on the cooling process of white dwarfs<sup>8,9</sup> have been already described. Here, we attempt to show that a galactic disk age of 15 Gyr cannot be excluded by the white dwarf observations if carbon and oxygen are immiscible in solid phase.

The luminosity function of white dwarf stars (the number of stars per unit volume as a function of their luminosity) increases monotonically with their absolute visual magnitude, and displays a well defined cutoff at magnitudes  $M_v > 16$  (refs 10, 11) ( $\log L/L_\odot \leq -4.5$ ). This behaviour can be understood in terms of the cooling of a hot and degenerate carbon–oxygen white dwarf core surrounded by a non-degenerate opaque envelope<sup>12</sup>. The total luminosity (photon plus neutrino) of the white dwarf

comes from the release of thermal energy (heat content plus latent heat when crystallization occurs), gravitational energy (which is negligible during late stages of cooling if the white dwarf core remains chemically homogeneous) and nuclear energy (hydrogen and helium burning at the surface during the early epochs). With these assumptions, it is possible to construct a theoretical luminosity function that recovers the observed luminosity function reasonably well in the range  $4 \geq \log(L/L_\odot) \geq -4$ , but predicts an excess of white dwarfs at lower luminosities if the commonly accepted age of the disk (15 Gyr) is adopted<sup>13</sup>. Thus, as the finite age of the galactic disk provides a natural cutoff to the luminosity, we can eliminate this excess of white dwarfs by assuming that the galactic disk is younger than previously thought. Using this approach and fitting observed and theoretical luminosity functions, Winget *et al.*<sup>2</sup> have estimated the age of the disk to be 9.3 Gyr, which implies that the Universe is 10.3 Gyr old.

The age derived this way depends critically on the assumption that the core of carbon–oxygen white dwarfs remains chemically homogeneous along the whole cooling process, even after solidification. But if we assume that carbon and oxygen separate in solid phase, the situation changes dramatically. Oxygen, being denser than carbon, sinks towards the centre, releasing gravitational energy. This process is dominant during the late stages of the cooling of white dwarfs. For instance, for a  $0.6M_\odot$  white dwarf, with equal mass fractions of carbon and oxygen, the gravitational energy released this way is  $e_{\text{grav}} \approx 10^{14} \text{ ergs g}^{-1}$ , which is larger than the latent heat of oxygen ( $\sim 5.2 \times 10^6 T \text{ ergs g}^{-1}$ , where  $T$  is the solidification temperature). Furthermore, this process happens when the central temperature is  $\sim 2 \times 10^6 \text{ K}$ , which corresponds to a luminosity of  $\sim \log L/L_\odot \approx -4$ .

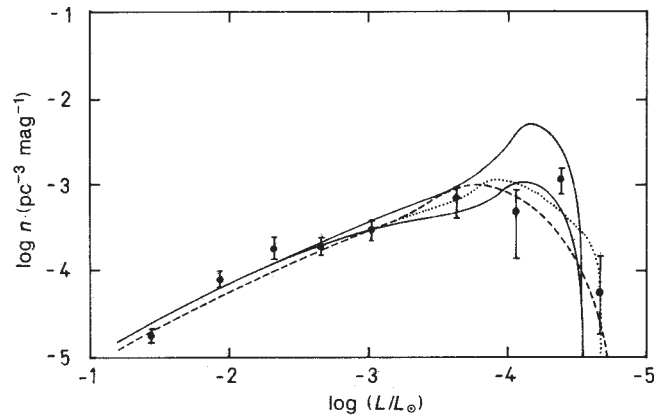
The confirmation of either carbon and oxygen miscibility or immiscibility in solid phase is a difficult problem as it requires the knowledge of the free energy of the completely ionized carbon–oxygen plasma in both liquid and solid phases. Using the Monte Carlo method, it is possible to compute the thermodynamic properties of the liquid phase. The free energy of the liquid can be well approximated by assuming ion-sphere charge averaging and an ideal entropy of mixing<sup>14</sup>. In solid phase the situation is more complex, as a given configuration has to be assumed for the alloy before computing the free energy. If we assume that ion-sphere charge averaging and ideal entropy of mixing are simultaneously valid for the solid, carbon and oxygen are miscible in solid phase. Stevenson<sup>3</sup> has pointed out, however, that both requirements cannot be satisfied simultaneously and adopted a model in which the solid alloy had an ideal entropy of mixing and the electrostatic energy of a random alloy. Because of the larger electrostatic energy of the alloy, the mixture tends to separate in solid phase. In the case of a completely ionized carbon–oxygen mixture, the phase diagram shows a eutectic for a carbon mass fraction  $X_c^e \approx 0.6$ , and total immiscibility in solid phase. The freezing temperature for different chemical compositions can be approximated on each side of the eutectic by:  $T/T_c = 0.925X_c^1 + 0.075$  ( $X_c^1 \geq X_c^e$ );  $T/T_c = 1.615 - 1.642X_c^1$  ( $X_c^1 \leq X_c^e$ ), where  $X_c^1$  is the carbon mass fraction in the liquid phase and  $T_c = 2.3 \times 10^4 \rho^{1/3}$  (where  $\rho$  is the density) is the freezing temperature of pure carbon. Of course, this phase diagram is an oversimplification and its theoretical confirmation needs more detailed calculations. Recently, Barrat *et al.*<sup>4</sup> have calculated the phase diagram of binary hard-sphere mixtures (which accounts for the geometrical packing aspects of solubility) as a function of  $\alpha$ , the ratio of sphere diameters. They have found that: (1) if  $0.94 < \alpha < 1$ , the phases are miscible in all proportions in both phases; (2) if  $0.92 < \alpha < 0.94$ , the phase diagram is azeotropic; (3) if  $0.85 < \alpha < 0.92$ , the phase diagram is eutectic with partial miscibility and (4) if  $\alpha < 0.85$ , there is total immiscibility in solid phase. Therefore, as in the context of the ion-sphere model, the ratio of diameters of carbon and oxygen is 0.908 (ref. 15), the existence of eutectic is favoured

by geometrical considerations. It is thus justified to look for the consequences of chemical separation due to crystallization in the two extreme cases: no chemical separation and total immiscibility.

When carbon and oxygen separate, an oxygen core surrounded by a carbon-rich mantle gradually forms. The gravitational contribution to the luminosity is proportional to the growing rate of the oxygen core ( $L_{\text{grav}} = e_{\text{grav}} m_{\text{ox}}$ ). The constant  $e_{\text{grav}}$  can be easily calculated either by building white dwarf models with oxygen cores of different masses and comparing their binding energies<sup>8</sup>, or by identifying the chemical differentiation with a change of the local internal energy<sup>9</sup>. Our calculations were made on the assumption that the white dwarf has uniform composition ( $X_{\text{C}} = X_{\text{O}} = 0.5$ ) at the beginning of crystallization. Table 1 shows the luminosity of a  $0.6 M_{\odot}$  carbon-oxygen white dwarf as a function of time obtained with our evolutionary code<sup>8,9</sup> for three cases: (1) no separation at all (total miscibility of carbon and oxygen in solid phase); (2) partial separation (the oxygen excess over the eutectic composition accumulates at the centre of the star and the mantle solidifies without further separation) when it has reached the eutectic composition and (3) total separation (once the eutectic has been reached, carbon and oxygen crystallize at the same time but independently; the denser 'oxygen flakes' continue to sediment over the oxygen core and the lighter 'carbon flakes' rise to lower-density regions, where they remelt). As we can see, if separation occurs, the cooling process is considerably slowed down. For instance, in case (2) luminosities lower than  $10^{-4.5} L_{\odot}$  are reached 1.6 Gyr later than in case (1). In case (3), the luminosity can be maintained during  $10^{10}$  years without changing by more than a factor of three.

The release of such an amount of energy not only slows down the cooling process, but can also produce a prominent bump in the luminosity function, in apparent contradiction with the observations. The question, then, is why do we not see it? There are two reasons: (1) old white dwarfs have a larger scale height over the galactic plane than the young ones (of course, this is neither supported or ruled out by the observations, but it is a general trend for all disk stars<sup>16,17</sup>) and (2) only rather massive main sequence stars ( $M_{\text{MS}} \geq 1.5 M_{\odot}$ ) have had time enough to produce the coolest white dwarfs that we observe.

To obtain a rough estimation of the age of the disk, we have constructed theoretical white-dwarf luminosity functions in the way described by Garcia-Berro *et al.*<sup>9</sup>, taking the age of the disk as a free parameter. The main assumptions are: (1) the average properties of white dwarfs can be represented by those of  $0.6 M_{\odot}$  ones; (2) the adopted IMF (initial mass function) is given by  $\varphi(m) \propto m^{-2.35}$  and the star formation rate is constant; the product of both functions has been normalized to give a luminosity function consistent with observations at  $\log(L/L_{\odot}) = -3$  (ref. 2); corresponding present white dwarf birthrate is  $0.25 \text{ yr}^{-1}$ ; (3) in view of the lack of observational data, the scale height of different white-dwarf generations can be approximated (it is reasonable to adopt the same approximations as for Mira variables<sup>17</sup>) by  $h(\tau) = 200 \text{ pc}$  for  $\tau < 5 \text{ Gyr}$ ,  $h(\tau) = 1,320 \text{ pc}$  for  $\tau > 12 \text{ Gyr}$  and  $h(\tau) = 160\tau - 600$  for  $5 < \tau < 12 \text{ Gyr}$  (where  $\tau$  is the



**Fig. 1** White-dwarf luminosity functions in the three cases shown in Table 1. Model 1, dashed line; model 2, dotted line; model 3, continuous lines (the upper continuous line has been obtained neglecting the dilution effect introduced by the scale height of white dwarfs). The first three curves are normalized to give the observed number of white dwarfs at  $\log(L/L_{\odot}) = -3$ . The normalization factor for the upper continuous line is the same as for the lower one. Observational data are from Winget *et al.*<sup>2</sup>

total age of the white dwarf, namely the cooling time plus the main sequence lifetime of the progenitor); (4) we have only considered white dwarfs fainter than  $10^{-1} L_{\odot}$  to avoid the complications of the early phases (neutrino losses and hydrogen or helium burning at the surface). The luminosity function can be written as

$$n(M_{\text{bol}}) = \left\langle \frac{1}{V(\tau)} \frac{dt_{\text{cool}}}{dM_{\text{bol}}} \nu_{\text{wd}}(t_{\text{d}} - t_{\text{cool}}) \right\rangle \Delta M_{\text{bol}} = 1$$

where  $t_{\text{d}}$  is the age of the disk,  $t_{\text{cool}}$  is the cooling time of a white dwarf,  $\nu_{\text{wd}}(t_{\text{d}} - t_{\text{cool}})$  is the number of white dwarfs formed per year at the time  $t_{\text{d}} - t_{\text{cool}}$ , and  $V(\tau) = 2Sh(\tau)$  is the relevant volume of the galactic disk ( $S$  being the disk surface).

Figure 1 shows the best fit between theoretical and observed luminosity functions for the three cases shown in Table 1, the fit being quite good in all cases. The ages derived in this way are 8.3, 9.6 and 15.0 Gyr respectively. Four points must be outlined: first, the existence of a eutectic does not necessarily imply total immiscibility in solid phase. A solid alloy could exist with a composition depending on the details of the phase diagram and cooling process. In this case, we would obtain an age between cases 1 and 2. Second, as we have already mentioned, the luminosity remains constant during the separation process, blurring the age-luminosity relationship. Anyway, even if we knew the exact phase diagram, we would need a very accurate luminosity function to elucidate the origin of the cutoff, namely to decide whether the short fall is due to the youth of the disk or to chemical fractionation. Third, the validity of our results strongly depends on the dilution factor introduced by the scale height of different white-dwarf generations. Figure 1 shows the luminosity function obtained assuming no dilution in order to show the importance of this effect. Fourth, the results strongly depend on the assumed chemical composition. If on the one hand oxygen is dominant and already stratified, as suggested by Mazzitelli and D'Antonna<sup>18</sup>, the effect of separation of carbon from oxygen is negligible. But on the other hand, carbon-oxygen white dwarfs contain  $^{22}\text{Ne}$  coming from the burning of  $^{14}\text{N}$ ; its abundance is rather low ( $\sim 1\%$  in mass), but, because of its high neutron/proton ratio, its deposition at the centre could provide an amount of energy equivalent to that of oxygen.

We conclude that the gravitational release of energy due to

**Table 1** Ages of the models (Gyr)

| $-\log(L/L_{\odot})$ | Model 1 | Model 2 | Model 3 |
|----------------------|---------|---------|---------|
| 2.0                  | 0.11    | 0.11    | 0.11    |
| 2.5                  | 0.31    | 0.31    | 0.31    |
| 3.0                  | 0.71    | 0.71    | 0.71    |
| 3.5                  | 1.66    | 1.66    | 1.66    |
| 3.75                 | 2.81    | 2.41    | 2.41    |
| 4.0                  | 4.26    | 4.16    | 4.16    |
| 4.25                 | 5.91    | 7.26    | 9.41    |
| 4.5                  | 7.81    | 9.96    | 22.88   |



chemical fractionation of carbon and oxygen during crystallization (if indeed it occurs) is a major effect not ruled out by the observations. Thus we believe that before using the age of white dwarfs as a measure of the age of the disk, it is necessary to elucidate the behaviour of high-density mixtures at the onset of solidification, as well as the relative mass fraction of carbon and oxygen in white dwarfs and their distribution above the galactic plane.

Received 15 December 1987; accepted 8 April 1988.

1. Schmidt, M. *Astrophys. J.* **129**, 243–258 (1959).
2. Winget, D. E. *et al. Astrophys. J.* **315**, L77–L81 (1987).
3. Stevenson, D. J. *J. Phys.*, Suppl. No. 3 **41**, C2/61–64 (1980).
4. Barrat, J. L., Baus, M. & Hansen, J. P. *Phys. Rev. Lett.* **56**, 1063–1065 (1986).
5. Canal, R., Isern, J. & Labay, J. *Astrophys. J.* **241**, L33–L36 (1980).
6. Isern, J., Labay, J., Hernanz, M. & Canal, R. *Astrophys. J.* **273**, 320–329 (1983).
7. Hernanz, M., Isern, J., Canal, R., Labay, J. & Mochkovitch, R. *Astrophys. J.* **324**, 331–344 (1988).
8. Mochkovitch, R. *Astr. Astrophys.* **122**, 212–218 (1983).
9. Garcia-Berro, E., Hernanz, M., Mochkovitch, R. & Isern, J. *Astr. Astrophys.* **193**, 141–147 (1988).
10. Liebert, J. W. *A. Rev. Astr. Astrophys.* **18**, 363–398 (1980).
11. Liebert, J. W., Dahn, C. C. & Sion, E. M. *The Nearby Stars and the Stellar Luminosity Function*, IAU Coll. 76 (eds Phillips, A. G. D. & Uggren, A. R.) 103–106 (Davis, Schenectady, 1983).
12. Mestel, L. *Mon. Not. R. astr. Soc.* **112**, 583–597 (1952).
13. Iben, I. I. & Tutukov, A. V. *Astrophys. J.* **282**, 615–630 (1980).
14. Hansen, J. P., Torrie, G. M. & Vioillefosse, J. P. *Phys. Rev. A* **16**, 2153–2168 (1977).
15. Ichimaru, S., Iyetomi, H. & Tanaka, S. *Phys. Rep.* **149**, 191–205 (1987).
16. Mayor, M. *Astr. Astrophys.* **32**, 321–327 (1974).
17. Wyatt, S. P. & Cahn, J. H. *Astrophys. J.* **275**, 225–239 (1983).
18. Mazzitelli, J. & D'Antonna, F. *Astrophys. J.* **311**, 762–773 (1986).

## Neutron star disk formation from supernova fall-back and possible observational consequences

F. Curtis Michel

Space Physics and Astronomy Department, Rice University, Houston, Texas 77251, USA

We have examined a simple model for gravitational fall-back of supernova ejecta to estimate the effect of delayed fall-back on pulsar activity from the central neutron star. Direct dimensional arguments suggest that a Kepler disk will be created with a mass of the order of  $10^{-5} M_{\odot}$  for the probable initial magnetic fields and rotation rates. It has already been argued that such a disk would be required for pulsar action. The disk expected to be formed by fall-back is, at first, too distant to favour pulsar action. We therefore predict a neutron star disk that may be detectable when the nebula around SN1987A clears, although the same model also suggests that the pulsar action might not at first be taking place. Because this disk should be composed of the same sort of shocked and highly processed matter that presently maintains the supernova light curve through radioactive heating, the disk should be hot and luminous, thereby providing a point power-law optical source in the nebula. Thus, observational possibilities include: (1) observation of thermal radiation from the disk, (2) observation of X-rays from a transient episode of accretion on to the neutron star, and (3) a delayed turn-on of the pulsar.

Let us take as a simple supernova model a sphere of matter that is set into expansion with a velocity directly proportional to distance from the centre. As long as density declines with distance from the centre, a sufficiently large velocity gradient leads to all of the mass beyond some value  $M_0$  being ejected to infinity. For matter with velocity less than the escape velocity ( $V_0$ ), the matter will move out and then fall back with a characteristic time  $T$ , where

$$(V_0^2 - V^2)/V_0^2 = (T_0/T)^{2/3} \quad (1)$$

and  $T_0$  is the free-fall collapse time to form  $M_0$  from matter at rest. If  $V = \alpha r$  (where  $\alpha = \text{constant}$ ), the left-hand side can be written  $(r_0^2 - r^2)/r^2$ . For roughly constant density, however,  $M \approx kr^3$ , where  $M$  is the mass interior to  $r$  at the time of the explosion, and approximately  $M_0 - M \approx 3kr_0^2(r_0 - r)$  for  $r$  not too different from  $r_0$ . Thus, within the same approximation

$$(T_0/T)^{2/3} = \frac{2}{3}\beta(M_0 - M)/M_0 \quad (2)$$

and we see that, despite the approximations,  $M \rightarrow M_0$  within the time of a few  $T_0$ . The factor of  $\frac{2}{3}$  is somewhat gratuitous, but differentiating then gives

$$\dot{M} = \beta(M_0/T_0)(T_0/T)^{5/3} \quad (3)$$

where  $\beta$  is a dimensionless number that for the moment is taken to be unity. Equation (3) yields an estimate of the backfall rate at late times. Although the neutron star reaches its asymptotic mass almost immediately, there is substantial mass waiting to fall in. For these estimates,  $T_0 \approx 1$  s and  $M_0 \approx 1.4 M_{\odot} \approx 3 \times 10^{33}$  g are the order-of-magnitude values, which from equation 2 suggest that even after a year there is  $1.5 \times 10^{-5} M_{\odot}$  left gravitationally bound to the neutron star. When this mass eventually falls in, it is such a tiny addition to the neutron star that it is unlikely to be structurally important. But this small mass can have a profound effect on its eventual functioning as a pulsar.

Infall is environmentally important when the dynamic pressure of the infall equals the output radiation pressure from the rotating ( $\Omega$ ) magnetized ( $B$ ) neutron star. As pointed out earlier<sup>1</sup>, a surrounding medium in hydrostatic equilibrium would intrude all the way into the light cylinder unless the dynamic pressure there exceeds the local infall pressure. A more conservative criterion is that the total pulsar energy output ( $L$ ) equals the gravitational potential energy gained by the infalling material, or

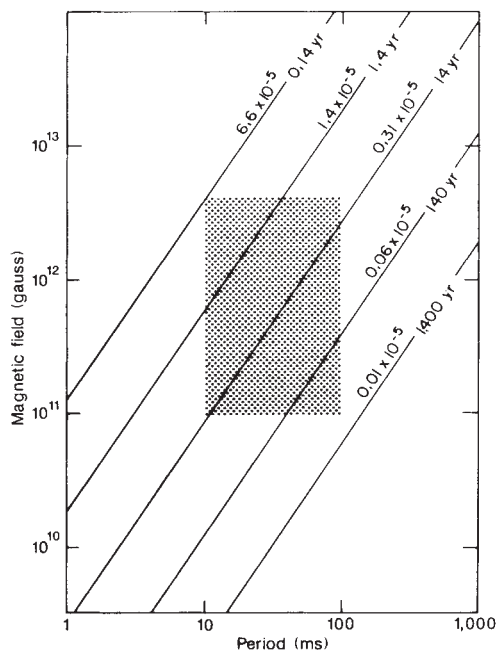
$$\dot{M} \frac{GM_0}{r_{lc}} = L \quad (4)$$

where  $r_{lc} = c/\Omega$ . Under such conditions, it is clear that the electromagnetic interaction can force the infalling matter into rotation. If we take nominal numbers<sup>2</sup> of  $\Omega = 600 \text{ rad s}^{-1}$ ,  $B = 4 \times 10^{12}$  G, hence  $L = 5.6 \times 10^{40} \text{ erg s}^{-1}$  and  $r_{lc} = 5 \times 10^7$  cm, we find a critical infall rate of  $\dot{M} = 1.4 \times 10^{22} \text{ g s}^{-1}$  and  $T = 6.3 \times 10^6$  s, or 0.14 yr. Given this one value, all others follow by scaling, using  $L = B^2 \Omega^4$  to estimate the pulsar power output (Fig. 1).

Our straightforward interpretation is that for the first 0.14 yr, such a pulsar would be smothered by infall<sup>3</sup>, and probably be essentially an accreting X-ray source. At later times, the radiation would be able to break out and stagnate the infall. It seems entirely reasonable to suppose that the stagnated infall is whipped into corotation, but at the same time it must collapse into a disk owing to the tidal forces that act even on a disk with a Keplerian rotation rate. Thus infall continues but is rechannelled into the formation of a disk. The disk in turn subtends a smaller cross-section to the radiation pressure and becomes dominated by centrifugal and gravitational forces. The mass of the disk is therefore essentially the mass of the remaining material which is in turn given from the time at which the infall ceases to overwhelm the pulsar,  $T = 6.3 \times 10^6$  s corresponding to  $M_0 - M = 6.3 \times 10^{-5} M_{\odot}$ . In this way we can assess the disk mass corresponding to pulsar initial conditions as also shown in Fig. 1.

For the probable initial conditions (shaded area), we see that disk masses of the order of  $10^{-5} M_{\odot}$  are expected. Coincidentally, such masses correspond to those proposed in a disk model for pulsar action<sup>4</sup>, and the relative insensitivity to initial conditions is consistent with the central hypothesis that pulsar disks are ubiquitous.

The above arguments lead to a prediction: SN1987A harbours a newly formed disk. Owing to the tiny size of the system,



**Fig. 1** Critical infall times as a function of neutron star magnetic field and spin period (likely initial values shown shaded). At longer times, the system should be forming a disk with an eventual mass as shown (order of  $10^{-5}$  solar masses). Lighter disks are formed for slow neutron stars with weak magnetic fields because they simply accrete the infall for most of its duration.

$10^{-5} M_{\odot}$  has a relatively large mean density, and the disk should essentially be a tidally compressed slice of white-dwarf material<sup>5</sup>. Because this material was at or near the same mass cut that ejected  $0.07 M_{\odot}$  of  $^{56}\text{Ni}$ , it should be intensely radioactive. The opacity for conversion of the decay products into heat is essentially unity, unlike the nebular opacity which should drop as  $\sim (1 \text{ yr}/T)^2$ . In addition, as the nebula expands, resolution of a central source reduces background by another factor proportional to  $T^2$ . Thus we expect the disk to be seen as a hot point source, which may be supplementally heated by interaction with the neutron-star electromagnetic fields that formed the disk in the first place.

How big is the disk and what would be the spectrum? If we invert equation 4 and write  $L(r) = L_0 r_{\text{ic}}^2 / r^2$ , we have  $r = L_0 r_{\text{ic}}^2 / GM_0 \dot{M}$ , and we can estimate where the mass is being deposited as a function of the deposition rate, and hence the residual mass to be deposited, because  $\dot{M} \approx M^{5/2}$  from equations 2 and 3, thus  $M(>r) \approx r^{-2/5}$  where  $M(>r)$  is now the disk mass beyond  $r$ . The surface density is therefore locally  $\sigma(r) \approx r^{-12/5}$ . In principle, this would correspond to a disk that is opaque beginning at an inner edge at  $\sim 5 \times 10^7 \text{ cm}$  (10 ms pulsar) and  $\sigma \approx 3 \times 10^{12} \text{ g cm}^{-2}$  and continues out to transparency ( $\sigma \approx 1 \text{ g cm}^{-2}$ ) at  $\sim 5 \times 10^{12} \text{ cm}$ . Note that it takes time to create these extended outer portions of the disk because  $r \approx T^{5/3}$ . If the heating is from radioactivity uniformly distributed in the disk, then  $dL/dr \approx \omega^4 d(\text{area}) \approx \omega^4 r$  but  $dL/dr = dM(>r)/dr \approx r^{-7/5}$  for radioactive heating, giving  $\omega^4 \approx r^{-12/5}$  or  $\omega \approx r^{-3/5}$  as the thermal profile. The flux density is  $S(\omega) d\omega \approx dL$ , hence  $S(\omega) = (dL/dr)(dr/d\omega) \approx \omega^{-1/3}$ , which gives a power-law spectrum in between the thermal cut-offs at the inner and outer disk temperatures. With these same scalings,  $\omega \approx 1/T$ , so this temperature range depends sensitively on when the disk first began to form. With such a spectrum, most of the energy is from the inner regions of the disk and most of the photons are from the outer regions.

Insofar as pulsar action goes, it should be noted that such infall-created disks are much larger in scale than those proposed

by Michel and Dessler<sup>4</sup>, where the assumption was made that the disk was immediately left over from the explosion and therefore occupied the probably allowed regions during an explosion, namely the region between the corotation distance and the pre-collapse core radius. Most importantly, having a disk with the inner edge at the light cylinder and extending beyond is almost irrelevant to the pulsar electrodynamics as envisioned in this model for pulsar activity<sup>5</sup>. Consequently, a certain period of time is expected to pass during which the disk evolves and the inner edge advances toward the neutron star before pulsar action is initiated. It is difficult to predict this time interval, but it is possible that shock excitation by the non-axisymmetric rotating magnetosphere of the neutron star produces sufficient effective viscosity<sup>6-8</sup> to promote inward advance of the disk on a reasonably short time scale ( $\sim$  years). Because such a mechanism is such a strong function of distance, it might at first drive an accretion and thinning of the near disk (another or continued episode as an X-ray source) before radiofrequency pulsations can be seen. Presence of a pulsar should nevertheless be revealed by non-radioactive powering of the nebula<sup>2</sup>.

These considerations follow entirely for dimensional analysis and scaling. Clearly, numerical estimates from detailed models of supernova explosions can be made to evaluate the dimensionless correction factor  $\beta$ . One obvious correction is for the density at the infall mass cut  $M_0$ , which for  $\beta = 1$  is taken to be equal to the average density interior. For a presupernova core with a reasonable polytropic index and mass twice that of the neutron star, this density ratio reduces  $\beta$  by about a factor of 3. The effect on the disk mass and smothering times is to multiply them by  $\beta^{3/5}$ . The velocity profile is more uncertain, with  $\beta = d(\log V)/d(\log r)$ , but there is no *a priori* expectation that such a factor would differ from unity by a large factor. These dimensional arguments do not, of course, preclude the possibility that the infalling mass can be accreted or ejected rapidly enough to avoid disk formation, a possibility undoubtedly attractive to proponents of pulsar action from isolated neutron stars. Nor can we be assured that the disk is not eroded by other processes (such as heating from the hot neutron star to produce a corona with subsequent evaporation). It should be noted, however, that the disk itself is in a deep potential well that would require (for a crab-like spin rate) about 10 MeV per particle for these processed nuclei to be removed from the system. Even if all the intercepted ( $\sim 1\%$ ) radiation from the star were diverted to expel particles, it would require thousands of years to remove the disk which is longer than the neutron star would be hot. More likely, the heated electrons should just transfer any energy from Compton scattering to the disk where it will be reradiated. In any case, the possible existence of a disk opens new and unexpected observational opportunities, some of which are: (1) the disk itself could be a significant source of 'thermal' radiation (albeit not a black-body spectrum), (2) accretion on to the neutron star could produce a pulsating non-binary X-ray source, and (3) a delayed turn-on of pulsar activity might be observed.

I am particularly indebted to Stanford Woosley who argued in a private letter that continuing infall might stifle pulsar activity, which directly stimulated these estimates (which were actually undertaken with the expectation that continuing infall would prove to be unimportant). Donald Clayton had earlier pointed out to me that such a disk should be quite radioactive. This work was supported in part by the NSF.

Received 16 February; accepted 3 May 1988.

1. Michel, F. C. *Astrophys. J.* **312**, 271-275 (1987).
2. Michel, F. C., Kennel, C. F. & Fowler, W. A. *Science* **238**, 938-940 (1987).
3. Manchester, R. N. & Taylor, J. H. in *Pulsars* 78-80, (Freeman, San Francisco, 1977).
4. Michel, F. C. & Dessler, A. J. *Astrophys. J.* **251**, 654-664 (1981).
5. Michel, F. C. *Astrophys. J.* **266**, 188-200 (1983).
6. Michel, F. C. *Astrophys. J.* **279**, 807-813 (1984).
7. Sawada, K., Matsuda, T. & Hachisu, I. *Mon. Not. R. astr. Soc.* **219**, 75-88 (1986).
8. Spruit, M. C. *Astr. Astrophys.* (Submitted).



## Frequency stability of solar oscillations

S. M. Jefferies\*, P. L. Pallé†, H. B. van der Raay‡, C. Régulo† & T. Roca Cortés†

\* Bartol Research Foundation, Delaware University, Delaware 19711, USA

† Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain

‡ Physics Department, Birmingham University, Birmingham B15 2TT, UK

Changes in the internal structure of the Sun over the 11-year magnetic activity cycle could be reflected in the eigenfrequencies of the acoustic p-modes. The first tentative experimental evidence was presented in 1984<sup>1</sup> and subsequently an analysis of ACRIM solar intensity data<sup>2</sup> suggested a decrease of frequencies of the 5-min solar p-modes between 1980 and 1984 of  $\sim 0.4 \mu\text{Hz}$ . Recently<sup>3-6</sup> further experimental data have provided conflicting results; frequency increases, decreases and stability have all been reported.

Extensive data obtained at the Observatorio del Teide (Izaña, Tenerife) from 1977 to 1985 are analysed to determine the lifetimes and frequencies of the  $l_0$  and  $l_1$  p-modes. The importance of the lifetime of the mode in determining the frequency is considered. From an analysis of the phase and amplitude behaviour of individual modes, a measure of the lifetime is obtained and the frequency of the mode over this lifetime is then found. Several such determinations yield the mean yearly frequencies which are then compared. No conclusive evidence for any frequency variation over the period of observation is found.

A possible cause for the conflicting results is to be found in a consideration of the lifetimes of the p-modes considered. By analogy with a stochastically excited oscillator, a particular mode in the Sun may be driven off its natural resonance by the continuous driving force. The period between such changes or re-excitations, constitutes the lifetime of the mode. Assuming the lifetime to be  $\tau$ , then upon re-excitation a particular mode could differ in frequency from its earlier value by  $\delta f \approx 1/\tau$ . Taking a value of  $\tau = 50$  days yields  $\delta f = 0.2 \mu\text{Hz}$  with correspondingly larger values if the lifetimes were shorter. Hence great care must be taken in deducing frequency variations of the solar p-modes if only discrete data sets of  $\sim 50$ -day duration from one year are compared with those from another.

The continuous existence of a particular mode may be found from a study of the amplitude and phase behaviour of a specified frequency over the period of observation. This may be shown by plotting the amplitude and phase of a specific mode as a function of time. Discontinuities in the amplitude, other than the amplitude modulation caused by rotational effects, may be linked to re-excitation of a particular mode. In general, this would be correlated with a phase discontinuity and possibly a change in the slope of the phase plot if a slightly different frequency were excited. Thus a study of the phase and amplitude

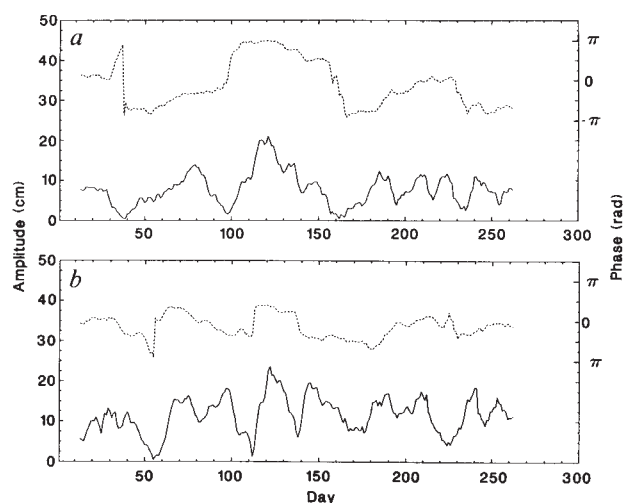


Fig. 1 The phase and amplitude variation of *a*, the  $l=0$ ,  $n=19$  line and *b*, the  $l=1$ ,  $n=20$  line as observed in 1985.

behaviour of individual modes will indicate the times at which re-excitation occurs and hence lead to a direct determination of the lifetime. The specific frequency during that lifetime may then be found and its uncertainty determined either from the lifetime or, if long data stretches ( $>200$  days) are available, the spread in frequencies resulting from several re-excitations will directly yield the mean frequency and its uncertainty. The mean yearly frequencies of each mode and their uncertainties may then be compared over the solar cycle to determine if any significant frequency changes occur.

The data used in this analysis were obtained from full disk resonant scattering determinations of the solar line-of-sight velocity. A potassium vapour cell was used to sample the 769.9-nm solar line. Each daily data set of  $\sim 900$  velocity determinations over an average 10-h observing period is corrected for instrumental response<sup>7</sup> and analysed to yield a set of residuals. The data available for analysis are summarized in Table 1.

The residuals are analysed by iterative sine-wave fitting with a frequency increment of  $0.1 \mu\text{Hz}$  to yield the mean amplitude and phase of each selected frequency over the data span considered. Then for a specific frequency, corresponding to a selected mode, the amplitude and phase variation over the entire year's data span can be obtained. As 1985 yielded the longest available data set, the results obtained will be illustrated by these data.

A conflict arises when aiming to determine the optimum length of data to be used to find the mean amplitude and phase. The satisfactory resolution of a particular mode from the other signals present in the data requires the longest possible data span consistent with the lifetime of the mode, yet to study the temporal variation of its amplitude and phase the highest resolution is obtained by the shortest data span. Spans of 15–30 days were considered and yielded similar results; as a compromise a span of 25 days was chosen for this analysis. Days 1–25 are used to determine the amplitude and phase of a particular frequency and then days 2–26 and so forth until the entire data span is scanned. The resulting spectra corresponding to a frequency of  $2,764.3 \mu\text{Hz}$  ( $l=0$ ,  $n=19$ ) are shown in Fig. 1*a*. The phase jumps at days 30, 95, 160 and 230 are well correlated with the amplitude variations and the different slopes of the phase plots show the changes in frequency when the oscillation is re-excited. The behaviour of the  $l=1$ ,  $n=20$  line during 1985 is shown in Fig. 1*b*. The phase jumps at days 55, 110, 140, 180 and 225 are again correlated with the major amplitude changes. The modulation of the amplitude signal during a phase coherent time interval is due to rotational effects. An analysis of this modulation leads to a determination of the rotational splitting<sup>8</sup>.

Table 1 Dates of observation used in this analysis

| Year | From     | To          | Span (days) | Data (days) |
|------|----------|-------------|-------------|-------------|
| 1977 | 12 July  | 25 August   | 45          | 37          |
| 1978 | 31 July  | 9 September | 41          | 25          |
| 1980 | 18 July  | 17 August   | 31          | 29          |
| 1981 | 29 May   | 25 August   | 88          | 80          |
| 1982 | 17 April | 6 September | 142         | 122         |
| 1983 | 10 May   | 2 September | 115         | 78          |
| 1984 | 17 April | 18 January  | 276         | 180         |
| 1985 | 27 March | 31 December | 279         | 210         |

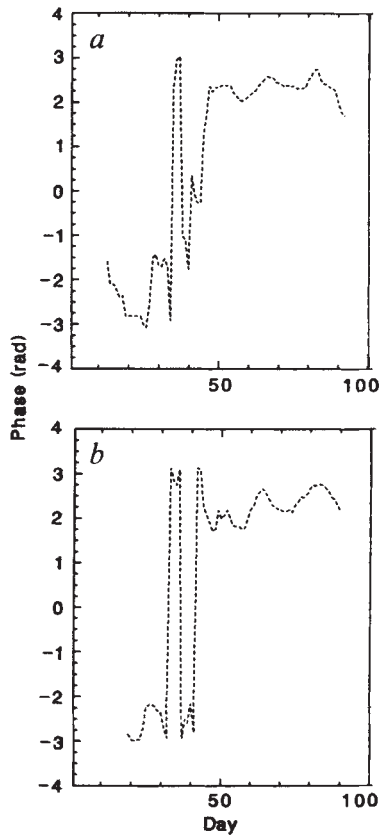


Fig. 2 A comparison of the phase variation of data from *a*, Izaña and *b*, Hawaii for the  $l=1$  line at  $3,098.7 \mu\text{Hz}$  obtained during 1983.

It might be thought that the observed phase jumps are caused merely by gaps in the data record, only 210 days of data were available in the 279-day data span. But during 1983 data were available from observations taken at Haleakala in Hawaii (8 June to 30 August an 84-day span with 74 days of data) and these when similarly analysed were compared with those taken at Izaña. The relevant phase plots for the dominant  $l=1$ ,  $n=21$  line at  $3,098.7 \mu\text{Hz}$  are shown in Fig. 2. The window functions of these two data sets being uncorrelated, the phase jump at day 30 can only be ascribed to a solar origin. Similar correlations are seen when other lines are considered and have been reported when 1981 data from the two sites were considered<sup>1</sup>.

The mean lifetimes deduced from this analysis for the  $l=0$  and  $l=1$  modes is shown for the 1985 data in Fig. 3. Linear fits to these data yield a lifetime of  $50 \pm 3$  days for  $l=0$ ,  $n=20$  ( $47 \pm 4$  days in 1984) with a slope of  $-1.5 \pm 0.5$  days per order ( $-1.0 \pm 0.7$  days per order in 1984) and  $45 \pm 3$  days for  $l=1$ ,  $n=20$  ( $48 \pm 2$  days in 1984) with a slope of  $-0.4 \pm 0.5$  days/order ( $-1.0 \pm 0.4$  days per order in 1984).

No correlation of the re-excitation times of specific modes of various orders was found for 1984 and 1985 data.

Having determined the specific days for which each particular mode exists, the power spectrum centred on that mode is obtained over its lifetime. Considering the  $l=1$ ,  $n=20$  mode shown in Fig. 1*b*, the power spectra for the time intervals day 55–110 and 140–180 are shown in Fig. 4*a* and *b*, respectively. The lines at  $2,963 \mu\text{Hz}$  are well resolved, the effects of rotational splitting are clearly evident and there is a discernible frequency shift between the two sets of data. The sidebands separated by  $11.57 \mu\text{Hz}$  from the true peaks result from the fact that only  $\leq 10$  h of data are available in every 24 h. By considering all five intervals defined by the phase jumps as shown in Fig. 1*b*, a mean frequency of  $2,963.3 \pm 0.2 \mu\text{Hz}$  is found. The error of  $0.2 \mu\text{Hz}$ , which is the standard deviation of the five frequency determina-

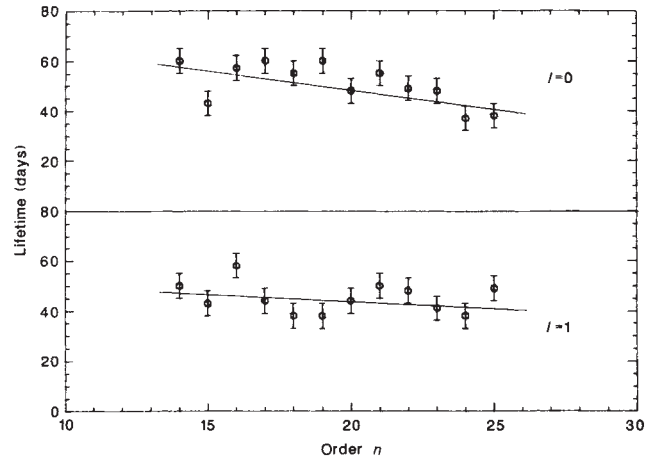


Fig. 3 The experimental lifetimes of the  $l_0$  and  $l_1$  modes deduced from the 1985 data. A mean decrease of  $\sim 1$  day per order is obtained from the fitted lines.

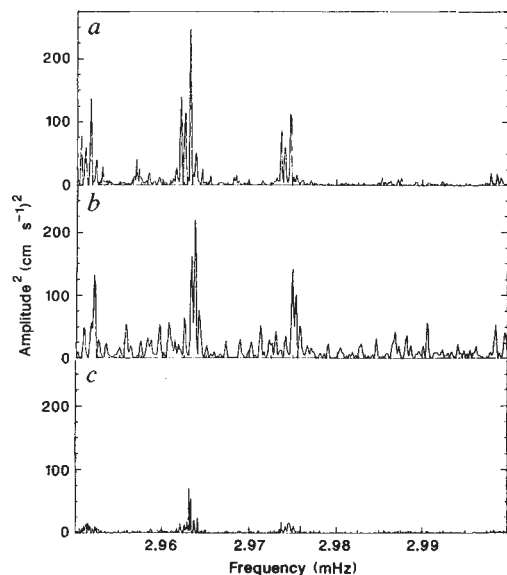


Fig. 4 Power spectrum for the  $l=1$ ,  $n=20$  line observed during 1985. *a*, Days 55–110; *b*, days 140–180 and *c*, days 1–279.

tions, is a further confirmation of the lifetime of the mode determined previously.

If, however, the whole year is considered as one continuous data string, considerable reduction in the apparent amplitude of the mode is found, as well as a somewhat broad, ill-defined spectrum. This is shown in Fig. 4*c* where the power spectrum is taken over days 1–279. Thus although longer data strings result in a reduction of noise because the signal is random, so too is the true solar oscillation signal reduced if data stretches longer than the lifetime are considered. Spectra such as that shown in Fig. 4*c* can lead to an underestimate of both the amplitude of a mode and its lifetime. As the individual frequency jumps associated with re-excitation of the mode are difficult to separate from the splitting caused by rotation, the overall broad feature not only suggests a short lifetime but can also lead to an incorrect determination of the mean frequency and the rotational splitting.

The mean frequencies of the  $l=0$  and  $l=1$  modes for  $n$  values from 14 to 26 are determined for each year where possible. A



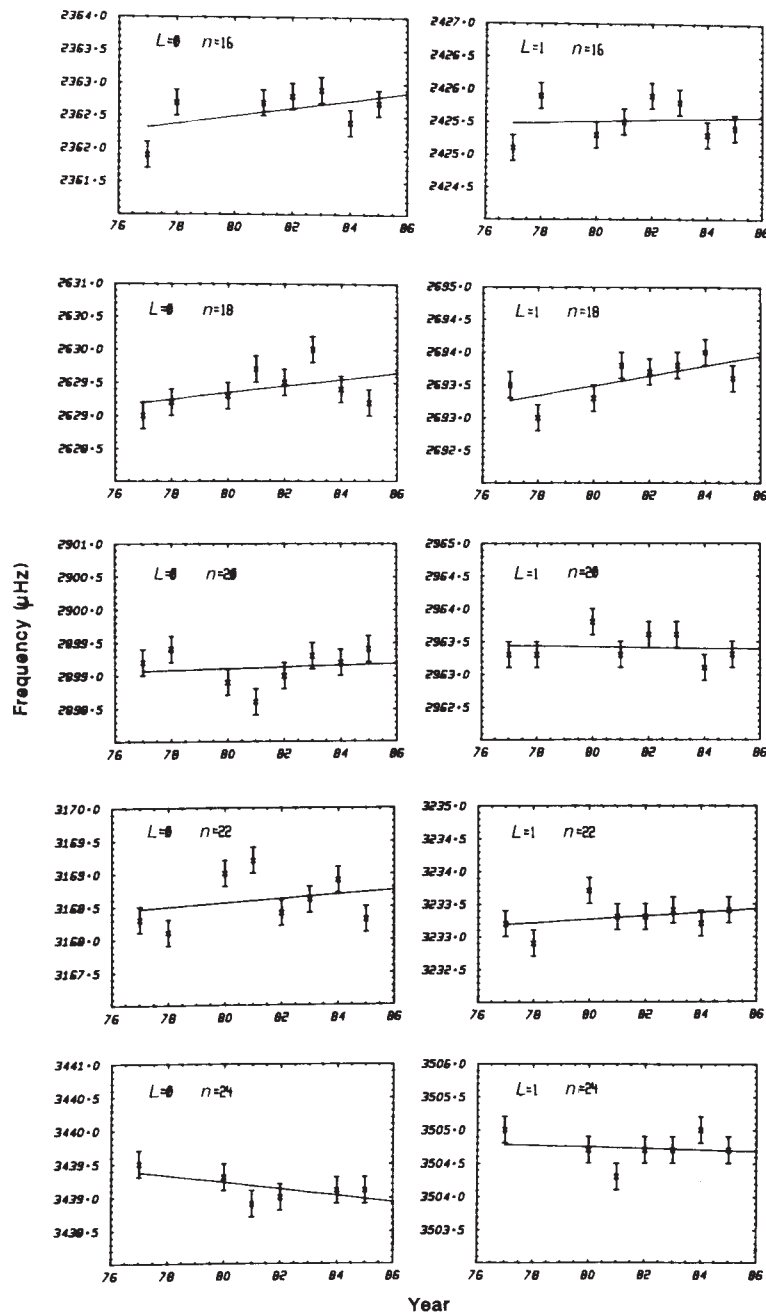


Fig. 5 The frequency variation of the  $l=0$  and  $l=1$  modes for various  $n$  values over the years 1977–85.

representative sample of the results obtained is shown in Fig. 5. The data are well fitted by straight lines and partial fits of data pre- and post-1980, show no consistent change in slope, thus indicating no significant correlations with the solar cycle. The slopes of the fitted lines, the frequency shift per year, are plotted in Fig. 6. The larger errors associated with the low-order and high-order values reflect the difficulty in determining the precise frequencies when the amplitudes of the signals become comparable with the noise level.

Both the  $l=0$  and  $l=1$  modes are consistent with no frequency shift from 1977 to 1985; the overall mean shift for the  $l=0$  mode being  $0.0 \pm 0.01 \mu\text{Hz yr}^{-1}$  and that for the  $l=1$  mode  $0.04 \pm 0.03 \mu\text{Hz yr}^{-1}$ . The corresponding values for 1980–85 are  $-0.01 \pm 0.02 \mu\text{Hz yr}^{-1}$  for the  $l=0$  mode and  $-0.01 \pm 0.02 \mu\text{Hz yr}^{-1}$  for the  $l=1$  mode.

In the analysis of solar oscillation data, the effects of the lifetimes of the modes should be considered when aiming to determine precise frequencies as these may vary from one excitation to the next. These effects have been directly measured by a determination of the mode lifetime from an analysis of the

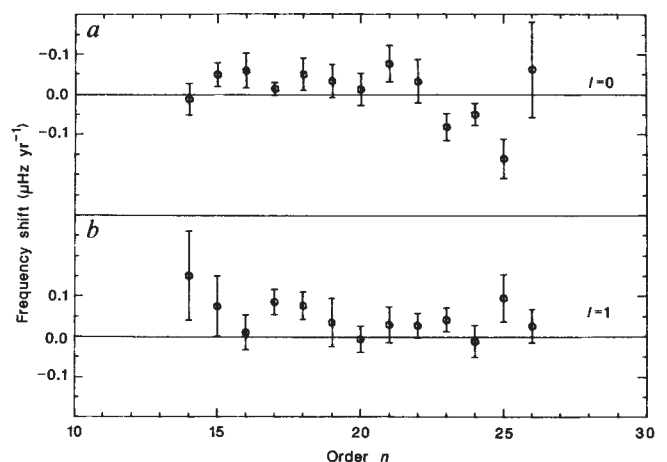


Fig. 6 The mean frequency shift over the years 1977–1985 for  $a$ ,  $l=0$  and  $b$ ,  $l=1$  modes.

phase and amplitude behaviour. The mean lifetimes are found to be  $\sim 50$  days at order 20 with a reduction of  $\sim 1$  day per order as the order of the mode increases.

The mean frequencies of the  $l=0$  and  $l=1$  modes were found for each year from 1977 to 1985. Considering the  $l=0$  modes with  $n=14$ –26, no net frequency shift was found. This should be contrasted with an earlier preliminary analysis of data from 1980 to 1984<sup>4</sup> which indicated a slight, although not statistically significant, positive shift. This analysis of the  $l=1$  modes confirms that no conclusive evidence for a frequency shift is found.

The assistance of all members of the Birmingham and I.A.C. solar oscillation groups is gratefully acknowledged, especially the support of the technical staff. This work was partly funded by SERC (UK) and CAICYT (Spain). Financial assistance from the British/Spanish Joint Research Programme (Acciones Integradas) is also gratefully acknowledged.

Received 21 December 1987; accepted 12 April 1988.

1. van der Raay, H. B. in *Theoretical Problems in Stellar Stability and Oscillations*, 215–219 (eds Noels, A. & Gabriel, M.) (University of Liege, 1984).
2. Woodard, M. F. & Noyes, R. W. *Nature* **318**, 449–450 (1985).
3. Woodard, M. F. & Noyes, R. W. in *Advances in Helio- and Asteroseismology* (eds Christensen-Dalsgaard, J. & Frandsen, S.) 197–200 (Reidel, Dordrecht, 1986).
4. Isaak, G. R., Jefferies, S. M., McLeod, C. P., New, R., van der Raay, H. B., Pallé, P. L., Régulo, C. & Roca Cortés, T. in *Advances in Helio- and Asteroseismology* (eds Christensen-Dalsgaard, J. & Frandsen, S.) (Reidel, Dordrecht, 1986).
5. Rhodes, E. J. Jr, Woodard, M. F., Cacciani, A., Tomczyk, S., Korzennik, S. G. & Ulrich, R. K. *Astrophys. J.* (in the press).
6. Fossat, E., Gelly, B., Grec, G. & Pomerantz, M. *Astron. Astrophys. Lett.* **177**, 47–48 (1987).
7. van der Raay, H. B., Pallé, P. L. & Roca Cortés, T. in *Seismology of the Sun and the Distant Stars* (ed. Gough, D. O.) 333–337 (Reidel, Dordrecht, 1986).
8. van der Raay, H. B., Pallé, P. L. & Roca Cortés, T. in *Seismology of the Sun and the Distant Stars* (ed. Gough, D. O.) 215–221 (Reidel, Dordrecht, 1986).

## The history and decay of a Mediterranean salt lens

Laurence Armi\*, Dave Hebert†, Neil Oakey‡, James Price§, Philip L. Richardson§, Thomas Rossby|| & Barry Ruddick†

\* Scripps Institution of Oceanography, La Jolla, California 92093, USA

† Department of Oceanography, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

‡ Department of Fisheries and Oceans, Bedford Institute of Oceanography, Dartmouth, Nova Scotia B2Y 4A2, Canada

§ Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

|| Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02882, USA

Subsurface coherent vortices called Meddies<sup>1</sup> are formed by the outflow of salty water from the Mediterranean Sea<sup>1,2</sup> into the North Atlantic. In October 1984 we began a study to observe the life history and large-scale changes of a Meddy by identifying a specific Meddy, and carefully mapping it and seeding it with Sofar floats<sup>3</sup>. (These neutrally buoyant floats are tracked acoustically and can be located aboard ship.) As this Meddy moved southward across the Madeira Abyssal plain, it was resurveyed three more times during a span of two years. Being able to find this same lens (100 km in diameter) on successive surveys was itself a unique achievement that allowed us to observe the Meddy evolution and to gain new insight into the importance of different mixing mechanisms that cause Meddy decay. We find evidence of mixing by at least three processes: (1) lateral mixing by the exchange of layers of water ('thermohaline intrusions')<sup>4,5</sup>, (2) vertical mixing at the underside of the lens by salt fingers<sup>6</sup> and (3) mixing by turbulence. Together these cause the net heat and salt anomalies to decay with

an e-folding time of about one year. Despite the mixing, the relative vorticity at the core remained constant for the first year and the Meddy retained its coherent shape over a two-year period.

The life of Meddies is of interest to oceanographers for at least two different reasons. One reason for tracking Meddies is to assess their role in the lateral dispersion of heat and salt<sup>2,7,8</sup>. The high-salinity Mediterranean outflow spreads into the North Atlantic to form a 'tongue' which can be observed from the coast of Portugal to well beyond the Mid-Atlantic Ridge. Also, many Meddies (formed from water from the same source) have been found in the Canary Basin to the south of this tongue<sup>2,7,9</sup>, and evidence for one near the Bahamas has been reported<sup>1</sup>. It is interesting to speculate that the salt tongue is due all or partially to decayed Meddies. To answer this we would need to know the number produced per year and their decay rate. By following one particular Meddy, the changes in lens properties can be observed, and the rates and mechanisms of exchange with the surrounding water can be inferred. A second reason for studying these lenses is that a Meddy might also be viewed as an isolated deep-water laboratory with its own dynamics, whose mixing behaviour observed over an extended period of time may be compared with the estimated changes from smaller-scale processes. To this end, the second visit included a complete survey of temperature and velocity microstructure using the deep-ocean microstructure profiler Epsonde<sup>10</sup>, to examine the role of turbulent mixing in the energetics of Meddy decay. In addition, there were velocity measurements from an acoustically tracked velocity profiler, Pegasus<sup>11</sup>, during the first and third

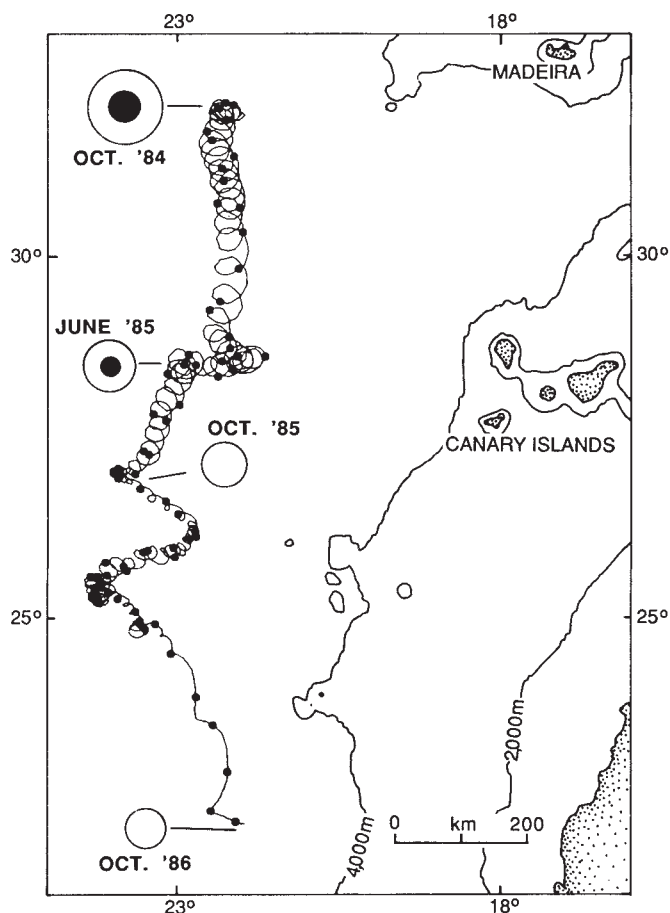


Fig. 1 Track of Sofar float EB128, which remained in the Meddy for nearly two years at a depth between 1,100 and 1,200 dbar. The dots indicate 10 day intervals. The 2,000 and 4,000-m isobaths are shown for reference. The circles indicate the size, structure, and position of the Meddy at each of the four surveys. Solid circles: Meddy Core Water. Open circles: Meddy Mixed Water.



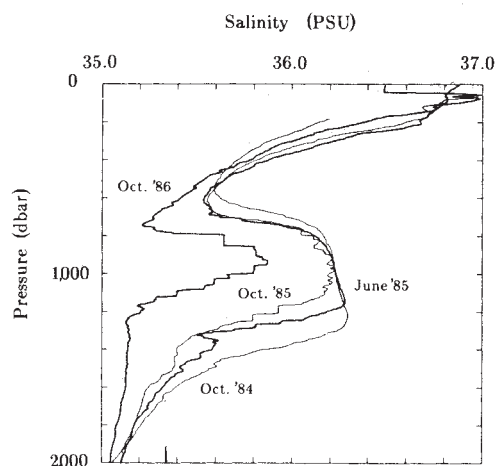


Fig. 2 Salinity profiles through the Meddy centre at each of the four surveys.

surveys, and from a mooring and expendable current profilers (XCPs) during the second survey, which allowed us to examine temporal changes in the vorticity as the Meddy decayed.

The two-year track of one of the Sofar floats implanted in the lens in October 1984 is shown in Fig. 1. The float trajectory consists of a slow, irregular translation due to the mean motion of the Meddy centre, plus a looping motion. The translation was mainly to the South, covering 1,100 km in two years. There is no clear scientific explanation why the Meddy moved southward rather than westward as suggested by Killworth<sup>12</sup>. The irregular motion would be consistent with the advection of the lens in the mesoscale eddy field. The looping motion is due to the anticyclonic motion about the Meddy centre which has a period of about six days. This period remained nearly constant throughout the experiment, except near the end, when both the period and the radius of looping increased.

In the discussion to follow, we define the anomalously warm, salty water of Mediterranean origin, smoothly stratified in both temperature and salinity as Meddy Core Water. In the initial survey this Meddy Core Water, which is anomalous in its temperature and salinity characteristics from the surrounding water in the Canary Basin (typically 11.9 °C and 36.2 PSU (practical salinity unit)) was found out to a distance 50 km from the lens centre before significant exchange with the background water was evident. Significant exchange was defined operationally as salinity profiles where the intrusions caused peak to peak fluctuations in salinity  $S_{pp} > 0.01$  PSU. The region beyond the core will be called Meddy mixed water (a mixture of core water and background water) and is delineated from the background by the operational criterion of  $S_{pp} > 0.05$  PSU. Vertical profiles of salinity through the centre of the Meddy from each of the surveys are shown in Fig. 2. When first observed in October 1984 the core extended from 650 to 1,400 dbar in total, with the region from 800 to 1,300 dbar being stable to double-diffusive processes. The core was thinner by June 1985, but still fairly smooth in vertical profile, showing no major effect of intrusions or mixing with the background water at the centre. The salinity of the core water was the same as in the previous survey: any apparent change in salinity is due to isopycnal displacement. By October 1985 intrusions reached the centre, as evidenced by the presence of salinity inversions. We refer to these signatures of Meddy mixed water as intrusions, but we know little of their horizontal structure. Some closely spaced conductivity-temperature-depth (CTD) profiles in the first three surveys indicate coherence of these inversions over distances of several kilometres. Although the Meddy Core Water was replaced by Meddy Mixed Water to the centre in about a year, the salinity anomaly was still  $\sim 0.6$  PSU, very little less than the salinity anomaly of the original Meddy Core Water. By the following year (October 1986) the

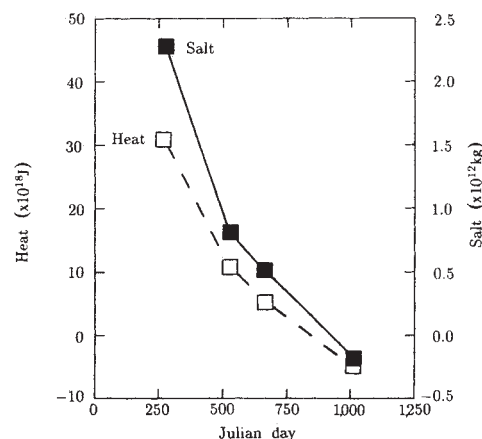


Fig. 3 Vertically and radially integrated salt (■) and heat (□) content for each of the surveys, computed relative to the background of the June 1985 survey. Day 1 is 1 January 1984.

inversions became much more step-like (suggestive of salt-fingering processes), and the salinity anomaly was much smaller. It was also apparent that the Meddy has drifted into fresher water, characteristic of 22° N.

The radial extent of the Meddy Core Water at each survey is shown as the filled circles in Fig. 1. Surrounding the core was a region of mixed water with a signature similar to the central profile of October 1985 of Fig. 2. This part of the lens contained many thermohaline inversions of up to 30 m vertical extent, and is represented in Fig. 1 by the area between the filled and open circles. By the second survey, the core water and surrounding mixed water decreased significantly in size. Did the Meddy slowly mix to this state, as suggested by the continued presence of the Sofar float in the lens, or were portions of it torn away by instabilities and turbulence? Although little turbulent mixing was observed in the Meddy core, turbulence and microstructure strongly suggestive of double-diffusive processes were very much in evidence at the Meddy periphery during the second survey<sup>10</sup>. By the third survey, the Meddy Core Water was entirely supplanted by Meddy Mixed Water. Inversions characteristic of intrusions were ubiquitous, but the salinity anomaly was greater than 0.6 PSU out to 19 km. This anomaly decreased markedly by the fourth survey. The total radius of the lens, including the mixed water region, decreased between the first and second survey, but did not decrease appreciably after that.

Figure 2 also documents the development of steps in the region underneath the Meddy core which is unstable to salt fingers. The steps have a salinity change of about 0.1 PSU, and become thicker and more sharply defined over time. Although Fig. 2 might suggest that we are seeing the evolution of individual steps, there is no evidence for this. The horizontal coherence of steps over distances much less than the lens radius was small even on a single survey. The apparent erosion from below is not simply an artefact of the vertical motion of isopycnals as the Meddy spins down; it takes place with respect to isopycnals as well. The precise thicknesses and salinities of the layers varied laterally, and from one survey to the next. Steps were observed at the base of the core for all surveys. Near the top of the Meddy, which is unstable to double-diffusive convection, steps were also observed but were not as pronounced as those below.

We computed the total excess salt content of the Meddy, relative to the background found at the June 1985 survey, as

$$\text{Total salt} = \int_0^{\max} 2\pi r dr \int \left( \frac{S(r, p) - S_{\text{ref}}(p)}{1,000} \right) g^{-1} dp \quad (1)$$

where the reference state  $S_{\text{ref}}$  is the salinity outside the Meddy during survey 2. The pressure integration extends only over the

depth range occupied by the Meddy at the central station to reduce the erroneous contribution of the changing background salinity. The excess heat content of the Meddy relative to the background was computed in a similar manner, and both are shown in Fig. 3. The values for the final survey are negative because of the choice of reference profile.

The initial excess salt content,  $2.5 \times 10^{12}$  kg, corresponds to the excess salinity contained in about 10 days of Mediterranean water flow through the Strait of Gibraltar at a volume transport of  $10^6 \text{ m}^3 \text{ s}^{-1}$  and an excess salinity of 2.5 PSU (ref. 13).

The Meddy has been envisaged as a lens of water, with salinity greater than 35.8 PSU, originally containing mostly core water and increasing amounts of mixed water as it decays. The Meddy lost substantial amounts of heat and salt. Salt was not conserved in our study because we chose to look only in the region of the strongest anomaly as described above. It is apparent that the mixing process carried salt and heat outside the well defined anticyclonic velocity core. The background velocity field may have carried this mixed water away, or otherwise the self-advection of the Meddy in the ambient potential vorticity gradient<sup>14</sup> carried the Meddy away from the mixed water. In this case, the Meddy would presumably leave a streaky convoluted 'salty trail' of mixed water as it moved. In fact, in the second survey, many 'blobs' of salty water with anomalies as high as 0.2 PSU were found to the east and southeast of the main Meddy. These may have been shed as the lens moved westward, preceding the time of the June 1985 survey.

The ratio of salt to heat loss was nearly constant throughout the two years. In terms of density effects, the two rates of loss were nearly equal, as expected for isopycnal processes. The rates of loss were twice as large ( $6.5 \times 10^4 \text{ kg s}^{-1}$  and  $9 \times 10^{11} \text{ W}$  respectively) in the first year, when an unadulterated core still existed, than they were in the second year, when the core was riddled with intrusions and the salinity anomaly was smaller. The loss rates correspond to an e-folding time of roughly one year, but are somewhat uncertain because of the changing background. The change in rate of loss may be related to the time at which intrusions reached the centre of the Meddy, or may simply be because the salinity anomaly was smaller in the second year.

On the basis of the above preliminary results, we put forth the following life history for this Meddy: (1) the Meddy remained a coherent, vortical structure for at least two years; (2) temperature and salinity inversions or intrusions filled the entire Meddy by the end of the first year as Meddy Core Water mixed laterally with background water to fill the entire Meddy volume with Meddy Mixed Water; (3) the mixed water outside the region of strong Meddy circulation was either left behind or swept away by background currents and eddies, causing a net loss of salt and heat from the Meddy; (4) during the second year, when the core was completely riddled with intrusions, the salinity anomaly at the centre decreased, and salt was lost more slowly than in the first year; (5) steps (a thermohaline staircase?) formed underneath the Meddy, where salt fingering is possible, and to a lesser degree above the Meddy where diffusive convection could occur; (6) the mixing processes did not appear to strongly transport angular momentum into or out of the core region. This is suggested by the roughly constant core vorticity, shown by the float track of Fig. 1, and by the supporting velocity profile studies. Whereas the salinity and vorticity in the core changed little over the first year, a net loss of angular momentum did occur as the core became smaller, with the removal of the mixed water from the vicinity of the Meddy; (7) turbulent mixing processes are active at the periphery of the Meddy, not only at the upper and lower boundary, but particularly in the mixed water region.

This informal collaborative experiment grew from the initial proposal of Armi and others joined with their respective contributions. The authors' names are assigned alphabetically. Dave Hebert was the scientist in charge of the reduction and analysis of the data from all four surveys.

Received 22 January; accepted 22 April 1988.

1. McDowell, S. E. & Rossby, H. T. *Science* **202**, 1085-1087 (1978).
2. Armi, L. & Zenk, W. *J. phys. Oceanogr.* **14**, 1560-1576 (1984).
3. Rossby, T., Voorhis, A. O. & Webb, D. *J. mar. Res.* **33**, 355-382 (1975).
4. Stern, M. E. *Deep Sea Res.* **14**, 747-753 (1967).
5. McIntyre, M. E. *Geophys. Fluid Dyn.* **1**, 19-57 (1970).
6. Turner, J. S. *Buoyancy Effects in Fluids* (Cambridge University Press, 1973).
7. Hebert, D. thesis, Dalhousie Univ. (1988).
8. McWilliams, J. C. *Rev. Geophys.* **23**, 165-182 (1984).
9. Käse, R. H. & Zenk, W. *J. phys. Oceanogr.* **17**, 158-163 (1987).
10. Oakey, N. S. *Oceans '87 Proc. Mar. Tech. Soc.* 316-321 (1987).
11. Spain, P. F., Dorson, D. L. & Rossby, H. T. *Deep Sea Res.* **28A**, 1553-1567 (1981).
12. Killworth, P. D. *J. phys. Oceanogr.* **16**, 709-716 (1986).
13. Gründlingh, M. L. *Meteor. Forsch.-Ergebnisse A/B* **23**, 15-46 (1981).
14. Rossby, H. T. *Eddies in Marine Science* (ed. Robinson, A. R.) (Springer, Berlin, 1983).

## A deep hydrological front between intermediate and deep-water masses in the glacial Indian Ocean

Nejib Kallel, Laurent D. Labeyrie  
Anne Juillet-Leclerc & Jean-Claude Duplessy

Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA,  
Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

Ocean circulation and climate are strongly interconnected. Under climatic conditions very different from those of today, deep and intermediate water circulation was subject to drastic changes<sup>1-7</sup>. For instance, during the last glacial maximum (LGM), deep ocean water was cooler than now by several degrees<sup>8</sup>. These temperature changes in the deep ocean were associated with striking variations in chemical characteristics of the intermediate and deep water masses of the Atlantic and Pacific Oceans<sup>3,8-13</sup>. Here we reconstruct the hydrological structure of the deep and intermediate water of the Indian Ocean during LGM (~18,000 yr BP). The carbon and oxygen isotope analyses of the benthic foraminifera genus *Cibicides* show that the water-column structure of the Indian Ocean during LGM was marked by the presence of a deep front separating intermediate and deep-water masses with very different characteristics. Intermediate-water mass temperatures and  $\delta^{13}\text{C}$  were similar to those of today. By contrast, the deep water was cooler than now by at least  $1.5^\circ\text{C}$ , more depleted in  $^{13}\text{C}$  and poorly oxygenated.

Under present conditions, the hydrological situation of the Indian Ocean is different from that of the other oceans. Intermediate water in the South Indian Ocean has a strong component of surface Antarctic water and is recognized by a salinity minimum to the north as far as  $10^\circ\text{S}$  (ref. 14). By contrast, in the North Indian Ocean (north of  $10^\circ\text{S}$ ), intermediate depths are mainly occupied by highly saline water from the Red Sea and the Arabo-Persian Gulf. Another contribution between  $10^\circ\text{S}$  and  $15^\circ\text{S}$  comes from the Pacific Ocean through the Indonesian Archipelago, and occupies depths between 500 and 1,200 m (ref. 15). Below 2,000 m, the Deep Indian Ocean Water originates mainly from the Southern Ocean<sup>16,17</sup>.

To determine the impact of the last glaciation on the Indian Ocean hydrological pattern, we have measured the foraminiferal  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  values in seventeen sediment cores from the two major basins of the North Indian Ocean, in the depth range from 1,250 m to 3,420 m (Table 1 and Fig. 1).

Kroopnick<sup>18</sup> has demonstrated that the distribution of  $^{13}\text{C}/^{12}\text{C}$  ratio of the  $\Sigma \text{CO}_2$  of sea water delineates the general distribution of water masses and that modern  $\delta^{13}\text{C}$  gradients in the deep ocean follow the net flow of the deep waters. Some species of benthic foraminifera (genus *Cibicides*) closely record the modern  $\delta^{13}\text{C}$  distribution in the world ocean, providing a good proxy for the reconstruction of the past  $\delta^{13}\text{C}$  gradients in the



Table 1 North Indian Ocean data

| Core      | Lat.    | Long.    | Depth (m) | GEOSECS temp. (°C) | Modern sea water $\delta^{18}\text{O}$ | Modern calculated Calcite $\delta^{18}\text{O}$ | Holocene benthic $\delta^{18}\text{O}$ | Analysed Holocene benthic foram | LGM <i>Cibicides</i> $\delta^{18}\text{O}$ | GEOSECS $\delta^{13}\text{C}$ | Holocene <i>Cibicides</i> $\delta^{13}\text{C}$ | LGM <i>Cibicides</i> $\delta^{13}\text{C}$ | Stratigraphic control for LGM | LGM level in core (cm) |
|-----------|---------|----------|-----------|--------------------|----------------------------------------|-------------------------------------------------|----------------------------------------|---------------------------------|--------------------------------------------|-------------------------------|-------------------------------------------------|--------------------------------------------|-------------------------------|------------------------|
| MD77-191  | 7.30 N  | 76.43 E  | 1,254     | 5.20               | -0.10                                  | 2.76                                            | 2.80                                   | <i>Uvigerina</i>                | 3.88                                       | 0.05                          | —                                               | 0.24                                       | A                             | 680                    |
| MD77-204  | 19.18 N | 58.26 E  | 1,430     | 5.25               | -0.12                                  | 2.72                                            | —                                      | —                               | 3.85                                       | -0.07                         | —                                               | 0.00                                       | A                             | 90                     |
| MD77-127  | 12.08 N | 75.9 E   | 1,610     | 3.80               | -0.12                                  | 3.14                                            | —                                      | —                               | 4.47                                       | -0.07                         | —                                               | -0.12                                      | B                             | 40                     |
| MD76-128  | 13.08 N | 73.19 E  | 1,712     | 3.60               | -0.12                                  | 3.16                                            | 3.13                                   | <i>Cibicides</i>                | 4.28                                       | -0.07                         | 0.09                                            | -0.15                                      | A                             | 210                    |
| MD76-125  | 8.35 N  | 75.20 E  | 1,878     | 2.96               | -0.19                                  | 3.27                                            | 3.26                                   | <i>Cibicides</i>                | 4.70                                       | 0.03                          | 0.29                                            | -0.08                                      | B                             | 120-130                |
| MD76-135  | 14.27 N | 59.31 E  | 1,895     | 3.30               | -0.14                                  | 3.32                                            | 3.05                                   | <i>Cibicides</i>                | 4.73                                       | -0.07                         | -0.14                                           | -0.07                                      | B                             | 220-240                |
| MD77-202  | 19.13 N | 60.41 E  | 2,427     | 2.12               | -0.24                                  | 3.44                                            | 3.46                                   | <i>Cibicides</i>                | 4.84                                       | -0.07                         | 0.17                                            | -0.39                                      | B                             | 265                    |
| MD77-203  | 20.42 N | 59.34 E  | 2,442     | 2.10               | -0.24                                  | 3.45                                            | —                                      | —                               | 4.80                                       | -0.07                         | —                                               | -0.38                                      | A                             | 380                    |
| SO28-57KL | 20.90 N | 63.12 E  | 3,422     | 1.50               | -0.27                                  | 3.59                                            | —                                      | —                               | 4.94                                       | 0.00                          | —                                               | -0.49                                      | (*)                           | —                      |
| MD77-176  | 14.50 N | 93.10 E  | 1,357     | 5.00               | -0.15                                  | 2.76                                            | —                                      | —                               | 3.87                                       | -0.01                         | —                                               | 0.04                                       | A                             | 565                    |
| MD77-169  | 10.13 N | 95.03 E  | 1,600     | 3.50               | -0.15                                  | 3.16                                            | —                                      | —                               | 4.33                                       | 0.00                          | —                                               | 0.00                                       | A                             | 190                    |
| MD77-171  | 11.46 N | 91.09 E  | 1,760     | 3.00               | -0.18                                  | 3.26                                            | —                                      | —                               | 4.38                                       | 0.00                          | —                                               | 0.03                                       | A                             | 280                    |
| RC11-147  | 19.04 S | 112.45 E | 1,953     | 2.34               | -0.23                                  | 3.39                                            | —                                      | —                               | 4.57                                       | 0.25                          | —                                               | 0.29                                       | (*)                           | —                      |
| MD77-179  | 18.30 N | 91.01 E  | 1,986     | 2.50               | -0.19                                  | 3.39                                            | —                                      | —                               | 4.90                                       | -0.02                         | —                                               | -0.23                                      | A                             | 40                     |
| MD77-181  | 17.40 N | 90.50 E  | 2,271     | 2.13               | -0.23                                  | 3.45                                            | —                                      | —                               | 4.97                                       | -0.02                         | —                                               | -0.74                                      | A                             | 130                    |
| MD77-182  | 16.10 N | 91.00 E  | 2,455     | 1.86               | -0.24                                  | 3.52                                            | —                                      | —                               | 5.09                                       | -0.02                         | —                                               | -1.08                                      | A                             | 60                     |
| MD77-183  | 15.17 N | 91.75 E  | 2,632     | 1.68               | -0.27                                  | 3.54                                            | —                                      | —                               | 4.98                                       | 0.02                          | —                                               | -1.09                                      | A                             | 130                    |

The data from the North Indian ocean show core location; modern sea water temperature<sup>30</sup> and  $\delta^{18}\text{O}$  (‰ vs PDB standard, ref. 29); estimates of modern  $\delta^{18}\text{O}$  of calcite in equilibrium with sea water (calculated using Shackleton relationship<sup>21</sup>);  $\delta^{18}\text{O}$  of the benthic foraminifera *Uvigerina* (core MD77-191) and *Cibicides* (corrected by +0.64‰) from the upper Holocene (upper Holocene isotopic values are from the tops of the cores MD 77-191, MD76-125 and MD76-135, the level 10 cm of the core MD76-128 and the level 32 cm of the core MD77-202);  $\delta^{18}\text{O}$  of *Cibicides* from LGM (corrected by +0.64‰); estimations of modern  $\delta^{13}\text{C}$  of the total dissolved  $\text{CO}_2$  (ref. 18) at the depth and location of the Indian deep sea core selection;  $\delta^{13}\text{C}$  of *Cibicides* from the upper Holocene;  $\delta^{13}\text{C}$  of *Cibicides* during last glacial maximum (LGM); stratigraphic control for LGM: A, continuous planktonic (*Globigerinoides ruber*)  $\delta^{18}\text{O}$  record (LGM was defined by the maximum in *Cibicides*  $\delta^{18}\text{O}$  values); B, continuous *Cibicides*  $\delta^{18}\text{O}$  record; \* data from ref. 35; LGM level in each core.

ocean<sup>2,19</sup>. The  $\delta^{18}\text{O}$  variations measured in foraminiferal shells reflect both the global variations of the ocean water  $\delta^{18}\text{O}$  due to continental ice volume changes and the isotope fractionation between calcium carbonate and water, which depends on the temperature at which foraminifera have formed their tests<sup>20-21</sup>. Along deep-sea cores, oxygen isotopic composition ( $\delta^{18}\text{O}$ ) changes of benthic foraminifera due to ice volume change are globally synchronous within the mixing time of the ocean ( $\sim 1,000$  yr)<sup>22,23</sup>. In each core, LGM level was recognized by the most recent maximum in the benthic  $\delta^{18}\text{O}$  record.

Oxygen and carbon isotope variations of the benthic foraminifera *Cibicides* have been measured on a VG 602 D mass spectrometer with a standard deviation of 0.07‰. Measured  $\delta^{18}\text{O}$  values of *Cibicides* are corrected by +0.64‰ to account for the deviation of these species from oxygen isotopic equilibrium<sup>2</sup>. Results are reported in Tables 1 and 2.

The  $\delta^{13}\text{C}$  values measured from Holocene *Cibicides* in four core tops (Table 1) show a +0.15‰ mean isotopic difference compared to modern values, estimated from GEOSECS water-column data<sup>18</sup>. This difference is barely significant, taking into account its 0.13‰ mean standard deviation. As the core top reference is missing for the other analysed cores, we have chosen to use GEOSECS data for modern conditions<sup>18</sup>. By so doing, we ignore local variability of the  $^{13}\text{C}$  distribution within the studied regions.

The  $\delta^{13}\text{C}$  values of glacial *Cibicides* in the eastern Indian Ocean are reported in Table 1 and Fig. 2b. A very strong and sharp discontinuity was separating the high  $\delta^{13}\text{C}$  values ( $\sim 0\text{‰}$ ) of Intermediate Water to a depth of  $\sim 2,000$  m from negative  $\delta^{13}\text{C}$  values (around  $-1\text{‰}$ ) deeper than 2,300 m, in sharp contrast to the modern situation where the vertical  $\delta^{13}\text{C}$  gradient in the North Indian Ocean is lower than 0.1‰ in this depth range (Table 1 and Fig. 2a and g). This originates from a decrease of the  $\delta^{13}\text{C}$  values of Bengal Gulf Deep Water by  $\sim -1\text{‰}$  during LGM (Fig. 2c). Previous studies have shown that  $\delta^{13}\text{C}$  changes of benthic foraminifera are due to changes in both the mean total dissolved  $\text{CO}_2$  content in the world ocean and the ocean circulation<sup>2,5,24,25</sup>. For the last deglaciation, change of the mean carbon isotopic composition of the total dissolved  $\text{CO}_2$  in the global ocean accounts for +0.32‰ (ref. 3). The small  $\delta^{13}\text{C}$  variation for Intermediate Water (Fig. 2c), compared with

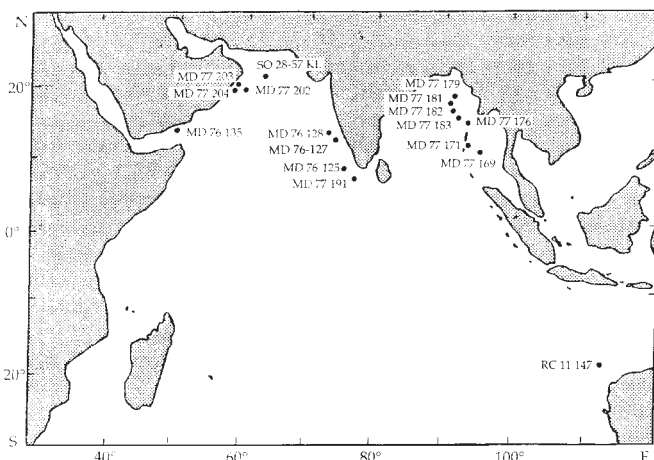


Fig. 1 Map showing core locations.

this global  $\delta^{13}\text{C}$  signal, indicates an enhanced glacial ventilation of this water mass, whereas the large  $\delta^{13}\text{C}$  change for Deep Indian Water indicates a very weak ventilation of the deep Indian Ocean during LGM. These results are in agreement with the already observed strong decrease in the ventilation of deep water in the glacial southern Ocean<sup>3,12</sup>. The depleted oxygen conditions in the glacial eastern North Indian Deep Water would be responsible for the absence of *Cibicides* in sediment cores deeper than 2,600 m, this genus being usually linked with well oxygenated sea water<sup>26</sup>.

*Cibicides*  $\delta^{13}\text{C}$  values of glacial Intermediate Water in the western and eastern sectors of the North Indian Ocean are comparable (Fig. 2b and h). Yet the Deep Water in the Arabian Sea had lower  $\delta^{13}\text{C}$  values than the Intermediate Water, but higher than in the Gulf of Bengal. Such asymmetry exists, although smaller, in the modern northern Indian Ocean because the passages connecting the basins of the western side are shallower and narrower than those of the eastern side. As a consequence, the hydrological characteristics of the western

**Table 2** Benthic isotope data surrounding the Last Glacial Maximum

| MD77-171   |                  |               |            |                         |               | MD76-128   |                         |               |
|------------|------------------|---------------|------------|-------------------------|---------------|------------|-------------------------|---------------|
| Level (cm) | $\delta^{18}$    | $\delta^{13}$ |            |                         |               | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |
|            | <i>Cibicides</i> |               |            |                         |               |            | <i>C. wuellerstorfi</i> |               |
| 260        | 3.18             | 0.18          | 110        | 4.18                    | -0.13         | 170        | 3.43                    | -0.03         |
| 270        | 3.29             | 0.21          | 120        | 4.29                    | -0.35         | 180        | 3.58                    | -0.11         |
| <b>280</b> | <b>3.74</b>      | <b>0.03</b>   | <b>130</b> | <b>4.33</b>             | <b>-0.74</b>  | 190        | 3.59                    | -0.13         |
| 290        | 3.52             | 0.10          | 142        | 4.26                    | -0.24         | <b>210</b> | <b>3.64</b>             | <b>-0.15</b>  |
| 300        | 3.73             | 0.01          | MD77-183   |                         |               | 220        | 3.54                    | -0.13         |
| 310        | 3.74             | -0.03         | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
| 320        | 3.64             | 0.04          |            | <i>Cibicides</i>        |               |            |                         |               |
| MD77-182   |                  |               | <b>130</b> | <b>4.35</b>             | <b>-1.03</b>  |            |                         |               |
| Level (cm) | $\delta^{18}$    | $\delta^{13}$ | <b>130</b> | <b>4.34</b>             | <b>-1.15</b>  |            |                         |               |
|            | <i>Cibicides</i> |               | 140        | 4.12                    | -0.16         |            |                         |               |
| 50         | 4.42             | -0.83         | 150        | 4.21                    | -0.51         |            |                         |               |
| <b>60</b>  | <b>4.45</b>      | <b>-1.08</b>  | 162        | 4.42                    | -0.41         |            |                         |               |
| 70         | 4.40             | -0.69         | MD77-176   |                         |               |            |                         |               |
| 80         | 4.21             | -0.29         | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
| 90         | 4.24             | -0.67         |            | <i>Cibicides</i>        |               |            |                         |               |
| MD77-203   |                  |               | 555        | 2.82                    | -0.04         |            |                         |               |
| Level (cm) | $\delta^{18}$    | $\delta^{13}$ | <b>565</b> | <b>3.23</b>             | <b>0.04</b>   |            |                         |               |
|            | <i>Cibicides</i> |               | 585        | 3.11                    | 0.02          |            |                         |               |
| 240        | 3.37             | -0.04         | 595        | 1.85                    | 0.26          |            |                         |               |
| 250        | 3.52             | -0.16         | 595        | 2.27                    | 0.12          |            |                         |               |
| 260        | 3.64             | -0.13         | MD77-179   |                         |               |            |                         |               |
| 270        | 3.72             | -0.16         | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
| 280        | 3.80             | -0.03         |            | <i>Cibicides</i>        |               |            |                         |               |
| 290        | 3.68             | -0.09         | 20         | 3.41                    | 0.08          |            |                         |               |
| 300        | 3.88             | -0.22         | 30         | 4.14                    | -0.24         |            |                         |               |
| 310        | 4.06             | -0.39         | <b>40</b>  | <b>4.26</b>             | <b>-0.23</b>  |            |                         |               |
| 320        | 3.98             | -0.28         | 50         | 3.90                    | -0.30         |            |                         |               |
| 330        | 3.96             | -0.32         | 60         | 4.05                    | -0.36         |            |                         |               |
| 340        | 4.08             | -0.49         | 70         | 4.10                    | -0.26         |            |                         |               |
| 350        | 4.08             | -0.27         | 80         | 4.01                    | -0.28         |            |                         |               |
| 360        | 4.06             | -0.45         | 90         | 3.77                    | -0.16         |            |                         |               |
| 370        | 4.08             | -0.39         | 100        | 3.90                    | -0.17         |            |                         |               |
| <b>380</b> | <b>4.16</b>      | <b>-0.38</b>  | MD76-127   |                         |               |            |                         |               |
| 390        | 3.51             | -0.23         | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
| 400        | 3.86             | -0.41         |            | <i>C. wuellerstorfi</i> |               |            |                         |               |
| 410        | 3.94             | -0.39         | 20         | 2.82                    | 0.21          |            |                         |               |
| 420        | 3.76             | -0.34         | 30         | 3.76                    | -0.06         |            |                         |               |
| 430        | 3.70             | -0.39         | <b>40</b>  | <b>3.83</b>             | <b>-0.12</b>  |            |                         |               |
| 440        | 3.91             | -0.45         | 50         | 3.71                    | -0.08         |            |                         |               |
| MD77-181   |                  |               | 60         | 3.48                    | -0.05         |            |                         |               |
| Level (cm) | $\delta^{18}$    | $\delta^{13}$ | 70         | 3.30                    | -0.10         |            |                         |               |
|            | <i>Cibicides</i> |               | 80         | 3.52                    | -0.28         |            |                         |               |
| 90         | 3.92             | -0.43         | 90         | 3.51                    | -0.23         |            |                         |               |
| 100        | 3.83             | -0.35         | 100        | 3.53                    | -0.51         |            |                         |               |
|            |                  |               | 110        | 2.79                    | -0.12         |            |                         |               |
|            |                  |               | MD77-191   |                         |               |            |                         |               |
|            |                  |               | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
|            |                  |               |            | <i>Cibicides</i>        |               |            |                         |               |
|            |                  |               | 650        | 2.37                    | 0.20          |            |                         |               |
|            |                  |               | 660        | 1.99                    | 0.17          |            |                         |               |
|            |                  |               | <b>680</b> | <b>3.24</b>             | <b>0.24</b>   |            |                         |               |
|            |                  |               | 690        | 3.20                    | 0.12          |            |                         |               |
|            |                  |               | 700        | 3.05                    | 0.00          |            |                         |               |
|            |                  |               | 710        | 2.58                    | 0.37          |            |                         |               |
|            |                  |               | 720        | 3.10                    | 0.00          |            |                         |               |
|            |                  |               | 730        | 2.15                    | 0.25          |            |                         |               |
|            |                  |               | MD77-169   |                         |               |            |                         |               |
|            |                  |               | Level (cm) | $\delta^{18}$           | $\delta^{13}$ |            |                         |               |
|            |                  |               |            | <i>Cibicides</i>        |               |            |                         |               |
|            |                  |               | 120        | 2.49                    | 0.20          |            |                         |               |
|            |                  |               | 140        | 3.18                    | 0.14          |            |                         |               |
|            |                  |               | 150        | 3.64                    | 0.00          |            |                         |               |
|            |                  |               | 160        | 3.55                    | 0.04          |            |                         |               |
|            |                  |               | 170        | 3.66                    | -0.11         |            |                         |               |
|            |                  |               | 180        | 3.57                    | -0.06         |            |                         |               |
|            |                  |               | <b>190</b> | <b>3.69</b>             | <b>0.00</b>   |            |                         |               |
|            |                  |               | 200        | 3.52                    | 0.01          |            |                         |               |
|            |                  |               | 210        | 3.19                    | -0.36         |            |                         |               |
|            |                  |               | 220        | 3.57                    | 0.00          |            |                         |               |

$\delta^{18}\text{O}$  and  $\delta^{13}\text{C}$  values of *Cibicides* around LGM for the different cores, with values at LGM written in bold characters (isotopic records of the core MD77-202 to be published in ref. 35 and for the cores MD76-135 and MD76-125 to be published in ref. 12).

deep water are altered by mixing with overlying Intermediate Water during its movement from the southern Ocean to the Arabian Sea. Deep Water in the eastern basins is more influenced by Antarctic Bottom Water and is thus colder than in the western basins<sup>14,27,28</sup>.

To examine the origin of the observed hydrological front between Intermediate and Deep Waters in the Indian Ocean, glacial  $\delta^{18}\text{O}$  values of the benthic foraminifera *Cibicides* were used as proxy indicators for temperature variations in the glacial Indian water column.

We have used the water  $\delta^{18}\text{O}$  values of the Scripps Institution of Oceanography shown by Birchfield<sup>29</sup> and temperatures<sup>30</sup> at the location and depth of each core to calculate the isotopic composition of calcite in equilibrium with modern sea water, using the Shackleton relationship<sup>21</sup> (Table 1). The calculated  $\delta^{18}\text{O}$  values are comparable to the measured  $\delta^{18}\text{O}$  of *Cibicides* (corrected for specific fractionation) from late Holocene core tops.

Calculated  $\delta^{18}\text{O}$  values for modern calcite and  $\delta^{18}\text{O}$  values for *Cibicides* during LGM from different deep sea cores in the two areas are reported versus depth in Fig. 2d, e, j and k respectively. The glacial  $\delta^{18}\text{O}$  profiles of *Cibicides* show a greater difference between Intermediate and Deep Water masses than the modern  $\delta^{18}\text{O}$  profiles, suggesting a sharp glacial hydrological gradient at the transition zone between them. This results in a discontinuity between Intermediate and Deep Waters similar to that marking  $\delta^{13}\text{C}$  profiles. The  $\delta^{18}\text{O}$  deep front is also more

pronounced in the eastern basins. This deep front could originate from temperature changes of either Intermediate or Deep Water masses.

Figure 2f, e show that the difference between measured  $\delta^{18}\text{O}$  values for LGM and those calculated for calcite in isotopic equilibrium with modern hydrological conditions is close to 1.1‰ for intermediate water. This is expected to account for the ice-sheet effect during the last deglaciation<sup>8,31</sup>.

Because of a lowering of sea level during LGM, Red Sea and Arabo-Persian Gulf contributions to Intermediate Water were probably very reduced or absent. Three sources are possible for the glacial Intermediate Water in the northern Indian Ocean: (1) dense water formed by increased evaporation and possible decreased temperatures of the surface water of the northern Arabian Sea<sup>32,33</sup>, (2) expansion to the north of 10° S of the Antarctic Intermediate Water (AAIW) in the absence of a northern local source and (3) increased flow of intermediate water from the Pacific Ocean through the Indonesian Archipelago. Support for this last possible origin comes from the analogy of the glacial  $\delta^{13}\text{C}$  (about +0.25‰) of the Intermediate Water of the western equatorial Pacific (ref. 3), the west of Australia (core RC11-147) and the west of Sri Lanka (core MD77-191), but a more complete coverage is required to determine unambiguously the origin of the glacial North Indian Intermediate Water. Such changes in the sources of intermediate water might have modified the vertical salinity gradient and correspondingly the water  $\delta^{18}\text{O}$  gradient. In the modern Indian Ocean the effect of



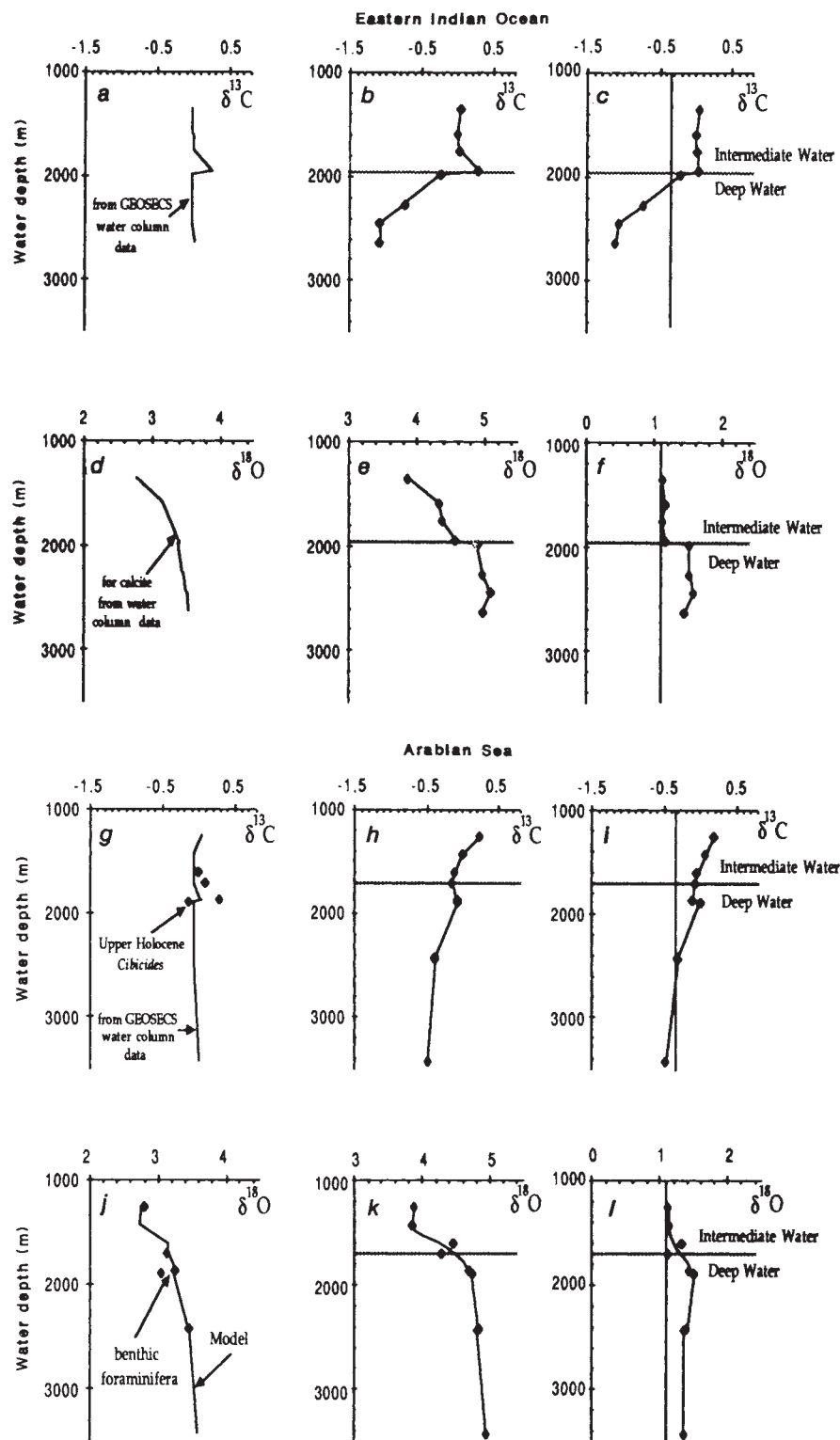


Fig. 2 Upper: Reconstruction of the isotopic vertical profiles in the eastern Indian Ocean  $\delta^{13}\text{C}$  profiles: *a*, for modern conditions<sup>18</sup> (the positive anomaly on this profile is due to the southern position of core RC11-147); *b*, for *Cibicides* during LGM; *c*,  $\delta^{13}\text{C}$  difference between LGM and today (the 0.32‰  $\delta^{13}\text{C}$  line represents the expected  $\delta^{13}\text{C}$  difference between LGM and today if there were no changes in the relative ventilation of water masses<sup>3</sup>).  $\delta^{18}\text{O}$  profiles: *d*, for modern calcite in oxygen isotopic equilibrium with sea water; *e*, for last glacial maximum *Cibicides*; *f*,  $\delta^{18}\text{O}$  difference between LGM and today (the 1.1‰  $\delta^{18}\text{O}$  line represents the expected  $\delta^{18}\text{O}$  difference between LGM and today, if there were no water temperature changes). Lower: Reconstruction of the isotopic vertical profiles in the Arabian Basin. *g*, Modern conditions (GEOSECS) ( $\delta^{13}\text{C}$  values of *Cibicides* from upper Holocene are also shown); *j*, for modern calcite in oxygen isotopic equilibrium with sea water (model) ( $\delta^{18}\text{O}$  values of benthic foraminifera from upper Holocene are also shown on this profile); *h*, *i*, *k* and *l* same as above (*b*, *c*, *e* and *f*), but for the Arabian Basin.

the highly saline Red Sea Intermediate Water source between 1,250 and 2,000 m accounts for a maximum increase of 0.1‰ in  $\delta^{18}\text{O}$  of the northern part of the Indian Ocean with respect to the southern one<sup>29</sup>. Modern Pacific Intermediate Water in the equatorial zone have  $\delta^{18}\text{O}$  characteristics similar to those of AAIW in the Indian Ocean<sup>29</sup>. Therefore the hypothesis that the  $\delta^{18}\text{O}$  gradient has remained constant, between glacial and modern conditions, may account for an overestimation of the Intermediate Water temperatures by only 0.4°C.

In the Deep Water, *Cibicides*  $\delta^{18}\text{O}$  values for LGM are heavier

than those expected to account for the continental ice volume effect by 0.41‰ in the eastern Indian Ocean and 0.31‰ in the Arabian Sea. If we assume no salinity effect on the Deep Indian Water  $\delta^{18}\text{O}$  other than that of the continental ice sheet, the measured foraminiferal  $\delta^{18}\text{O}$  variations would indicate that glacial Deep Water was cooler than now by ~1.5°C in the eastern Indian Ocean and 1.1°C in the Arabian Sea. The influence of the highly saline North Atlantic Deep Water was very weak or absent during LGM (refs 3, 10). But the maximum decrease in the Indian Deep Water  $\delta^{18}\text{O}$  due to the decrease in the North

Atlantic Deep Water influence has been estimated as 0.1‰ to maintain the source of AABW over the freezing point<sup>8</sup>. Thus the maximum possible change in the whole vertical  $\delta^{18}\text{O}$  gradient is 0.1‰.

These results indicate that the vertical salinity/ $\delta^{18}\text{O}$  gradient in the northern Indian Ocean during LGM was not so different from the modern one. The sharp deep front between Intermediate and Deep Water is thus best explained by the decrease in Deep Water temperature in the Indian Ocean and development of a deep thermocline. The increased hydrological contrast between Intermediate and Deep Waters in the Indian Ocean has amplified the asymmetry between eastern and western Basins induced by the bottom topography of this ocean.

Our results demonstrate that the glacial Indian Ocean hydrological structure was different from modern conditions. Intermediate Water was of similar temperature but more ventilated than now. Similar observations about the intermediate-water ventilation during LGM are made for the Atlantic and Pacific Oceans<sup>3,10,12,13</sup>, indicating the global nature of this phenomenon. By contrast, the deep world ocean was subject to drastic changes with a very weak ventilation and cooler temperatures. The description of the glacial ocean circulation has to take into account this strong stratification between intermediate and deep waters, which has, as described by Boyle<sup>34</sup>, a potential influence on the regulation of the atmospheric  $p\text{CO}_2$ .

We thank J. Agnignac, B. Le Coat and H. Leclaire for their role in the isotope analyses, J. Davies, M. Fontugne for discussion, and N. J. Shackleton and L. Keigwin for the review of the manuscript. This work was financed partly by the Centre National de la Recherche Scientifique and Commissariat à l'Energie Atomique and partly by the Institut National des Sciences de l'Univers (Programme National d'Etude de la Dynamique du Climat) and the Programme Interdisciplinaire de Recherche sur l'Environnement (GRECO  $\text{CO}_2$ ). The support of the Mission Scientifique des Terres Australes et Antarctiques Françaises has been especially important in providing the ship *Marion Dufresne* for the coring cruises OSIRIS.

Received 25 January; accepted 18 April 1988.

- Duplessy, J. C., Moyes, J. & Pujol, C. *Nature* **286**, 479–482 (1980).
- Duplessy, J. C. *et al. Quat. Res.* **21**, 225–243 (1984).
- Duplessy, J. C. *et al. Paleoceanography* (in the press).
- Shackleton, N. J., Imbrie, J. & Hall, M. A. *Earth planet. Sci. Lett.* **65**, 223–244 (1983).
- Duplessy, J. C. & Shackleton, N. J. *Nature* **316**, 500–507 (1985).
- Boyle, E. A. & Keigwin, L. D. *Science* **218**, 784–787 (1982).
- Boyle, E. A. & Keigwin, L. D. *Earth planet. Sci. Lett.* **76**, 135–150 (1986).
- Labeyrie, L. D., Duplessy, J. C. & Blanc, P. L. *Nature* **327**, 477–482 (1987).
- Labeyrie, L. D. & Duplessy, J. C. *Paleogeogr. Paleoclimatol. Paleoecol.* **50**, 217–240 (1985).
- Oppo, D. & Fairbanks, R. G. *Earth planet. Sci. Lett.* **86**, 1–15 (1987).
- Curry, W. B. & Lohmann, G. P. *Quat. Res.* **18**, 218–235 (1982).
- Curry, W. B., Duplessy, J. C., Labeyrie, L. D. & Shackleton, N. J. *Paleoceanography* (in the press).
- Boyle, E. A. & Keigwin, L. D. *Nature* **330**, 35–40 (1987).
- Wyrki, K. in *The biology of the Indian Ocean* (ed. Zeitzschel, B.) 18–36 (Springer, Berlin, 1973).
- Gordon, A. L. *J. geophys. Res.* **91**, 5037–5046 (1986).
- Mantyla, A. W. & Reid, J. L. *Deep Sea Res.* **30**, 805–833 (1983).
- Broecker, W. S., Takahashi, T., & Takahashi, T. *J. geophys. Res.* **90**, 6925–6939 (1985).
- Kroonnick, P. *Deep Sea Res.* **32**, 57–84 (1985).
- Zahn, R., Winn, K. & Sarnthein, M. *Paleoceanography* **1**, 27–42 (1986).
- Emiliani, C. *J. Geol.* **63**, 538–578 (1955).
- Shackleton, N. J. in *Variation du Climat au Cours du Pleistocène* (ed. Labeyrie, J.) 203–209 (CNRS Paris, 1974).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973).
- Duplessy, J. C. in *Climatic Change* (ed. Gribbin, J.) 46–67 (Cambridge University Press, 1978).
- Shackleton, N. J. in *The Fate of Fossil Fuel  $\text{CO}_2$  in the Oceans* (ed. Anderson, N. R. & Malahoff, A.) 401–427 (Plenum, New York, 1977).
- Broecker, W. S. *Prog. Oceanogr.* **11**, 151–197 (1982).
- Streeter, S. S. & Shackleton, N. J. *Science* **203**, 168–171 (1979).
- Warren, B. A. *Deep Sea Res.* **28A**, 759–788 (1981).
- Fioux, M., Schott, F. & Swallow, J. C. *Deep Sea Res.* **33**, 415–426 (1986).
- Birchfield, G. E. *Paleoceanography* **2**, 431–442 (1987).
- GEOSecs Indian Ocean Expedition Vol. 5 (eds Weiss, R. F., Broecker, W. S., Craig, H. & Spencer, D.) (IDOE NSF, Washington, DC, 1983).
- Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183–190 (1987).
- Duplessy, J. C. *Nature* **295**, 494–498 (1982).
- CLIMAP Project Members *Science* **191**, 1131–1137 (1976).
- Boyle, E. A. *Nature* **331**, 55–56 (1988).
- Sarnthein, M., Winn, K., Duplessy, J. C. & Fontugne, M. *Paleoceanography* (submitted).

## Climatic and $\text{CH}_4$ cycle implications of glacial–interglacial $\text{CH}_4$ change in the Vostok ice core

D. Raynaud\*, J. Chappellaz\*, J. M. Barnola\*,  
Y. S. Korotkevich† & C. Lorius\*

\* Laboratoire de Glaciologie et Géophysique de l'Environnement,  
BP 96 38402, St Martin d'Hères Cedex, France

† Arctic and Antarctic Research Institute, Beringa Street 38,  
199226 Leningrad, USSR

The atmospheric  $\text{CH}_4$  increase from  $\sim 0.7$  to  $1.68$  p.p.m.v. over about the past 300 years which has been documented from analysis of air trapped in ice cores<sup>1–4</sup> and from tropospheric measurements (see ref. 5 for example) is attributed to anthropogenic modifications of the  $\text{CH}_4$  cycle. The concern about this increase is due to the radiatively and chemically active nature of  $\text{CH}_4$ . Here we present strong evidence from analysis of the Vostok ice core, that  $\text{CH}_4$  concentrations increased from  $0.34$  to  $0.62$  p.p.m.v. between the end of the penultimate ice age and the following interglacial, about 160–120 kyr BP. This  $\text{CH}_4$  change may be explained by considering the effect of the climatic change on the  $\text{CH}_4$  cycle. Its contribution (including chemical feedback) to the global climatic warming is estimated to be about 25% of that due to  $\text{CO}_2$ .

The Vostok ice core provides a unique palaeoenvironmental record covering the past 160 kyr<sup>6–8</sup>. Our first aim has been to investigate whether atmospheric concentrations of  $\text{CH}_4$  were different during full glacial conditions from those prevailing in an interglacial period. Seven depth levels were selected: three from the penultimate glaciation ( $\sim 155$  kyr BP), three from the following interglacial ( $\sim 130$  kyr BP) and one in the transition in between. We chose this time period in preference to the last glacial maximum ( $\sim 18$  kyr ago)–Holocene period because of potential problems linked with the presence of fractures in the upper part of the Vostok core<sup>7</sup>.

Each ice sample ( $\sim 40$  g) is taken from the centre of the core and placed in a glass vessel sealed with vacuum silicone grease. The air surrounding the sample is evacuated, and the ice is melted. The meltwater is then slowly refrozen from the bottom. In this way, most of the air dissolved in the water is pushed out at the water–ice interface. Based on experimental tests, the efficiency of this extraction is estimated to be 99%. After refreezing the water, the gas is expanded in an extraction line and injected in a gas chromatograph (flame ionization detector).

Three successive injections from the same extracted air are performed and the system is calibrated using an air standard (Air Liquide) containing  $1.2 \pm 0.1$  p.p.m.v. of  $\text{CH}_4$  in a mixture of  $\text{N}_2$ ,  $\text{O}_2$  and  $\text{CO}_2$ . The analytical accuracy of each analysis ( $2\sigma$ ), evaluated from the standard deviation of the residuals corresponding to the calibration regression, is  $0.03 \pm 0.02$  p.p.m.v. Blank values are obtained by analysing standard gas in the same way as real samples, a cube of artificial bubble-free ice replacing the ice-core sample. Based on 12 tests, a correction that varies between about 0.04 and 0.1 p.p.m.v. has been applied to the results.

No trend appears between the three injections from the same sample, and our estimated overall accuracy (excluding the uncertainty due to the standard gas) is about 0.04 p.p.m.v. This value is consistent with the observed scattering ( $2\sigma$ ) of the results obtained from the same ice-core section (Table 1), suggesting that  $\text{CH}_4$  concentrations are homogeneous in a given core section.

Apart from the Vostok core, a section of another Antarctic core (D 57) has been analysed to compare the results obtained using our method with previously published data. The air occluded in the D 57 section was trapped around AD 1440, according to the dating based on volcanic horizons<sup>9</sup>. The mean



CH<sub>4</sub> concentration (Table 1) is  $0.68 \pm 0.04$  p.p.m.v., in very good agreement with previous results<sup>1-4</sup> that indicate a CH<sub>4</sub> concentration of  $\sim 0.7$  p.p.m.v. before the beginning of significant anthropogenic perturbation. The agreement is even better with results from three Byrd core samples<sup>3</sup> of about the same age, which give a mean CH<sub>4</sub> concentration of 0.67 p.p.m.v.

The results obtained on the seven Vostok core sections (27 samples analysed) and shown in Table 1, are plotted in Fig. 1 versus age, together with the isotopic temperature profile<sup>6</sup>. The age of the ice is from ref. 10 and the age of the air, which is younger, has been calculated according to the procedure described in ref. 7. The mean CH<sub>4</sub> concentration of the three interglacial levels (12 samples) is 0.62 p.p.m.v., with a maximum level value (0.65 p.p.m.v.) rather close to the fifteenth century value obtained from the D 57 ice core. We may also note a trend in the concentrations which parallels the temperature profile. The three levels (12 samples) covering the end of the penultimate glaciation give a mean CH<sub>4</sub> concentration of 0.34 p.p.m.v.. The level measured in the climatic transition shows an intermediate value of 0.46 p.p.m.v. The Vostok results thus indicate that the tropospheric CH<sub>4</sub> concentrations increased very significantly (by a factor 1.8) between the penultimate ice age and the following interglacial.

We now discuss whether the glacial-interglacial warming could itself have forced the CH<sub>4</sub> increase. We assume a similar warming for the last two glacial-interglacial transitions and use the estimate of  $4.5 \pm 1^\circ\text{C}$  (ref. 11) for such a global warming.

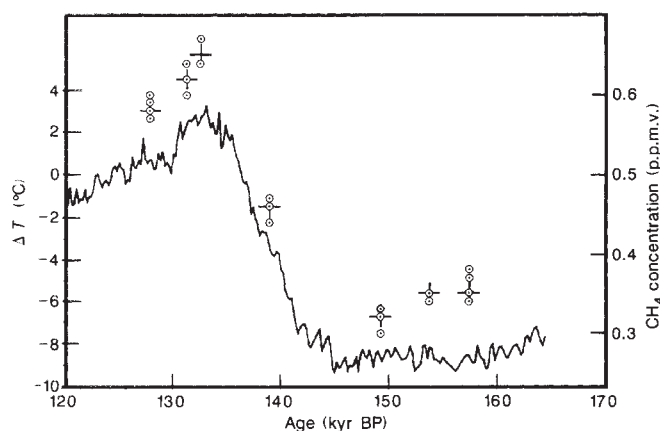
Oxidation of tropospheric CH<sub>4</sub> by OH ( $\text{CH}_4 + \text{OH} \rightarrow \text{CH}_3 + \text{H}_2\text{O}$ ) is the main identified sink for CH<sub>4</sub>. Because of the temperature dependence of the reaction rate coefficient, we expect an enhanced destruction of CH<sub>4</sub> by OH when the temperature increases. The strength of this sink may also increase because of an enhanced production of OH arising from the higher atmospheric H<sub>2</sub>O levels produced by warmer temperatures<sup>12,13</sup>. Typical results obtained from a coupled climate-chemical model<sup>12</sup> show that, for a mean surface temperature increase of  $3.1^\circ\text{C}$ , the enhanced sink strength leads to a tropospheric CH<sub>4</sub> decrease by 17%. Thus the effect of a  $4.5^\circ\text{C}$  warming on this sink during the penultimate deglaciation could be a decrease in the CH<sub>4</sub> concentrations by about 20%. There is evidence that unsaturated soils also act as a CH<sub>4</sub> sink, but it is unlikely that they are important in the global CH<sub>4</sub> budget (see ref. 14 for example). If our estimate of the change in the sink strength is correct, the global source strength must have increased by a factor of 2.3 to account for the observed increase in CH<sub>4</sub> concentration ( $\times 1.8$ ).

The sources of atmospheric CH<sub>4</sub> have been discussed by

**Table 1** CH<sub>4</sub> concentrations in D 57 and Vostok ice cores with depths and ages

| Depth (m)     | Age of the ice (yr BP) | Mean age of the gas | CH <sub>4</sub> concentration (p.p.m.v.) | Mean CH <sub>4</sub> concentration ( $\pm 2\sigma$ ) (p.p.m.v.) |
|---------------|------------------------|---------------------|------------------------------------------|-----------------------------------------------------------------|
| <b>D 57</b>   |                        |                     |                                          |                                                                 |
| 180.7         | 850                    | 510                 | 0.66, 0.72, 0.68<br>0.67, 0.68           | 0.68 (0.04)                                                     |
| <b>Vostok</b> |                        |                     |                                          |                                                                 |
| 1,789.2       | 130,080                | 127,800             | 0.58, 0.60, 0.57<br>0.58, 0.59, 0.58     | 0.58 (0.02)                                                     |
| 1,834.7       | 133,360                | 131,100             | 0.64, 0.60, 0.62                         | 0.62 (0.04)                                                     |
| 1,852.3       | 134,650                | 132,400             | 0.67, 0.64, 0.64                         | 0.65 (0.04)                                                     |
| 1,932.1       | 141,600                | 138,800             | 0.44, 0.46, 0.47                         | 0.46 (0.03)                                                     |
| 2,016.4       | 153,420                | 149,100             | 0.30, 0.32, 0.33                         | 0.32 (0.03)                                                     |
| 2,042.5       | 157,790                | 153,600             | 0.35, 0.34, 0.35                         | 0.35 (0.01)                                                     |
| 2,063.7       | 161,277                | 157,300             | 0.37, 0.34, 0.34<br>0.35, 0.38, 0.34     | 0.35 (0.04)                                                     |

Owing to the trapping process of the air by the ice, each level represents an average value of the age of the air—covering several hundred years in the case of Vostok<sup>7</sup> and  $\sim 50$  years for D 57.



**Fig. 1** Vostok ice core: CH<sub>4</sub> concentrations (+ indicate mean values) and isotope temperature profile (continuous line, from ref. 6) versus time.

several authors<sup>14</sup>. Based on the different estimates of CH<sub>4</sub>-emission rates and of their trends<sup>15</sup>, we can assume that the flux of methane evolved from bacterial processes in freshwater wetland areas was the dominant source prevailing before the anthropogenic perturbation. Although there are only limited CH<sub>4</sub> flux measurements at freshwater wetland sites, increasing temperature may lead to an exponential response of CH<sub>4</sub>-emission rate (as in ref. 12). Based on the CH<sub>4</sub>-emission rate response given in ref. 12 and a temperature change of  $4.5^\circ\text{C}$ , the interglacial source strength would be higher than the glacial by a factor of 1.7. The areal extent of such anaerobic environments should logically also have been larger during an interglacial than during a glacial period, because of wetter climate and more extended deglaciated area. A surface increase of  $\sim 35\%$ , combined with the temperature impact on the freshwater wetland source strength, would account for the CH<sub>4</sub> increase from the ice age to the interglacial. We are aware that one could argue in favour of other sources (termites, wildfires, wild ruminants), but our aim is to stress that climatically induced changes of a global CH<sub>4</sub> source, like freshwater wetlands, could well account for the observed CH<sub>4</sub> increase and that this increase could be entirely explained as being a consequence of the climatic change.

The climatic impact due to the observed glacial-interglacial CH<sub>4</sub> increase (0.34–0.62 p.p.m.v.) is an important issue. The corresponding direct radiative effect can be evaluated from radiative-convective models. The analytic expression provided by Hansen *et al.*<sup>16</sup>, which fits the results of a one-dimensional radiative-convective climate model<sup>17</sup> within about 1%, leads to a global surface equilibrium warming of  $\sim 0.08^\circ\text{C}$ . Furthermore, under present-day conditions, increasing CH<sub>4</sub> produces higher levels of tropospheric O<sub>3</sub>, another radiatively active gas, which in turn leads to a positive feedback on the surface temperature. Models taking this chemical feedback into account indicate an amplification of the direct CH<sub>4</sub> greenhouse effect by  $\sim 75\%$  (ref. 18) and  $58\%$  (ref. 19). The assumption that these amplification factors also applied at the glacial-interglacial transition would lead to a radiative equilibrium warming of  $0.13$ – $0.14^\circ\text{C}$  due to the CH<sub>4</sub> increase and its chemical feedback. This is about four times less than the equilibrium warming ( $0.5^\circ\text{C}$ ) obtained from the same model<sup>16</sup> for the CO<sub>2</sub> increase (from 195 to 270 p.p.m.v.) which occurred during the same time interval<sup>7</sup>.

According to these calculations, the combined net greenhouse effect due to CO<sub>2</sub> and CH<sub>4</sub> changes (and excluding climatic feedbacks) would have increased the global surface equilibrium temperature by  $\sim 0.6^\circ\text{C}$ . We are aware of the uncertainties linked with the evaluation of the different climatic feedbacks, but if we adopt a climatic feedback factor of 3.5, as suggested in ref. 20 by the Goddard Institute Space Studies three-dimensional climate model (yielding a sensitivity of  $\sim 4^\circ\text{C}$  for doubling CO<sub>2</sub>),

the global equilibrium warming arising from the  $\text{CO}_2$  and  $\text{CH}_4$  increases would be  $\sim 2.2^\circ\text{C}$  ( $0.5^\circ\text{C}$  for  $\text{CH}_4$  alone). Hence, the combined net climatic forcing of  $\text{CO}_2$  and  $\text{CH}_4$  would be about 50% of a glacial-interglacial warming of  $4.5^\circ\text{C}$ . If our suggestion that climate is initially forcing  $\text{CH}_4$  changes is correct, we can expect a  $\text{CH}_4$  contribution all along the temperature change observed during the last climatic cycle, and in particular in the isotope-temperature record obtained from the Vostok ice core. Further measurements of  $\text{CH}_4$  concentrations from this ice core will help to clarify this point.

These results have been obtained in the frame of a French-Soviet cooperation with logistic support of the US NSF (Division of Polar Programs). We thank J. Jouzel for providing the isotope data. Scientific discussions with P. Crutzen, A. Lacis, M. Legrand and H. Le Texier were most helpful. This work was supported by the Commission of the European Communities, CNRS/PIREN, TAAF and Soviet Antarctic Expeditions.

**Note added in proof:** After this paper was submitted, another paper<sup>21</sup> appeared sharing a very similar  $\text{CH}_4$  change between the last ice age and the Holocene. The two sets of data together strongly suggest that the observed  $\text{CH}_4$  increases as a general characteristic of the glacial-interglacial climatic transitions.

Received 29 February; accepted 5 May 1988.

1. Craig, H. & Chou, C. C. *Geophys. Res. Lett.* **9**, 1221-1224 (1982).
2. Khalil, M. A. K. & Rasmussen, R. A. *Chemosphere* **11**, 877-883 (1982).
3. Rasmussen, R. A. & Khalil, M. A. K. *J. geophys. Res.* **89**, 11599-11605 (1984).
4. Stauffer, B., Fischer, G., Neftel, A. & Oeschger, H. *Science* **229**, 1386-1388 (1985).
5. Blake, D. R. & Rowland, F. S. *Science* **239**, 1129-1131 (1988).
6. Jouzel, J. *et al. Nature* **329**, 403-408 (1987).
7. Barnola, J.-M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408-414 (1987).
8. Legrand, M. R., Lorius, C., Barkov, N. I. & Petrov, V. N. *Atmos. Environ.* **22**, 317-331 (1988).
9. Raynaud, D. & Barnola, J. M. *Nature* **315**, 309-311 (1985).
10. Lorius, C. *et al. Nature* **316**, 591-596 (1985).
11. Rind, D. & Peteet, D. *Quat. Res.* **24**, 1-22 (1985).
12. Hameed, S. & Cess, R. D. *Tellus* **35B**, 1-7 (1983).
13. Ramanathan, V. *et al. Rev. Geophys.* **25**, 1441-1482 (1987).
14. *World Meteorological Organization, WMO Report 16* (1985).
15. Bolle, H. J., Seiler, W. & Bolin, B. in *The Greenhouse Effect, Climatic Change and Ecosystems* (eds Bolin, B. *et al.*) 157-203 (Wiley, Chichester, 1986).
16. Hansen, J. *et al. J. geophys. Res.* submitted.
17. Lacis, A., Hansen, J., Lee, P., Mitchell, T. & Lebedeff, S. *Geophys. Res. Lett.* **8**, 1035-1038 (1981).
18. Owens, A. J. *et al. J. geophys. Res.* **90**, 2283-2311 (1985).
19. Brasseur, G. & De Rudder, A. *J. geophys. Res.* **92**, 10903-10920 (1987).
20. Hansen, J. *et al. Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 130-163 (American Geophysical Union, Washington, 1984).
21. Stauffer, B., Loehbronner, E., Oeschger, H. & Schwander, J. *Nature* **332**, 812-814 (1988).

## Plume-asthenosphere mixing beneath the Galapagos archipelago

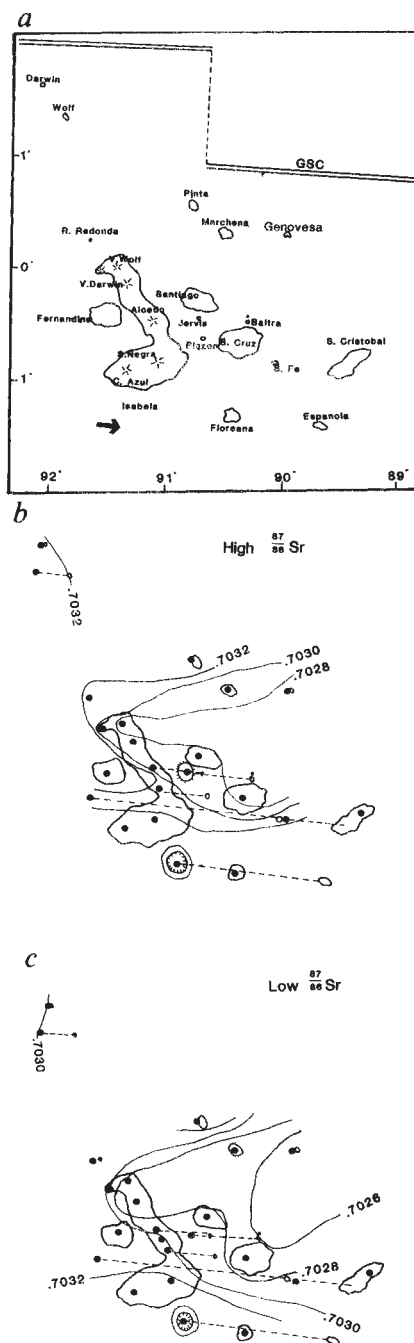
Dennis J. Geist\*†, William M. White‡§  
& Alexander R. McBirney||

\* Department of Geology and Geophysics, University of Wyoming, Laramie, Wyoming 82071, USA

‡ Max-Planck-Institut für Chemie, Mainz, 6500, FRG and College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

|| Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA

Although the Galapagos Islands have long been cited as an example of intraplate volcanism resulting from a mantle plume<sup>1,2</sup>, a number of features of the Galapagos appeared to conflict with a simple mantle-plume model. These include the non-linear arrangement of volcanoes, the non-time-transgressive nature of volcanism, and a tendency for the lavas to have more depleted trace-element and isotopic signatures towards the centre of the archipelago. The toroidal diapir model of Griffiths<sup>3,4</sup>, in which depleted asthenosphere is drawn into the centre of the rising diapir with the original



**Fig. 1** *a*, Map indicating locations discussed in text. Arrow shows absolute motion of the Nazca Plate<sup>19</sup>. *b*, Galapagos Islands contoured by the highest  $^{87}\text{Sr}/^{86}\text{Sr}$  ratios measured from each volcano. Position of each sample point has been adjusted (dashed lines) for its estimated age<sup>8,9,13,14,17,25,26</sup> and the velocity of the Nazca Plate<sup>19</sup>. *c*, Same as above, but for lowest measured  $^{87}\text{Sr}/^{86}\text{Sr}$  from each volcano.

diapir material preserved in a torus, appears to resolve these conflicts and provides an explanation for many of the enigmatic aspects of Galapagos volcanism.

The Galapagos Islands are thought to be the manifestation of a mantle diapir that lies immediately adjacent to the Galapagos spreading centre (GSC), an active mid-oceanic ridge (Fig. 1). Interaction of the hotspot with the ridge is apparent from anomalous geochemical features in GSC lava, such as high  $^{87}\text{Sr}/^{86}\text{Sr}$  and  $\text{La}/\text{Sm}$  and low  $^{143}\text{Nd}/^{144}\text{Nd}$  ratios<sup>5-7</sup>. But the Galapagos differ from more familiar hotspots, such as Hawaii, in several important respects. First, the islands do not form a single linear chain; instead they are aligned in two oblique

† Present address: Department of Geology, Hamilton College, Clinton, New York 13323, USA.

§ Present address: Department of Geological Sciences, Cornell University, Ithaca, New York 14853, USA.



**Table 1** Range of Sr and Nd isotopic ratios from individual Galapagos volcanoes

| Volcano                                      | $^{87}\text{Sr}/^{86}\text{Sr}$ | $^{143}\text{Nd}/^{144}\text{Nd}$ | <i>n</i> | Ref.    |
|----------------------------------------------|---------------------------------|-----------------------------------|----------|---------|
| Santa Cruz (young shield)                    | 0.70262–0.70274                 | 0.51306–0.51310                   | 6        | 13, 15* |
| Santa Cruz (old platform) Baltra, N. Seymour | 0.70286–0.70320                 | 0.51301–0.51304                   | 4        | 13, 15* |
| Pinta                                        | 0.70329–0.70338                 | 0.51289–0.51290                   | 6        | 15*, 17 |
| Genovesa                                     | 0.70259–0.70272                 | 0.51310–0.51313                   | 5        | 15*     |
| Alcedo                                       | 0.70309–0.70327                 | 0.51298–0.51300                   | 2        | 15*     |
| Ecuador                                      | 0.70297                         | 0.51301                           | 1        | 15*     |
| Sierra Negra                                 | 0.70300–0.70338                 | 0.51292–0.51293                   | 3        | 15*     |
| Cerro Azul                                   | 0.70332                         | 0.51293                           | 1        | 15*     |
| Darwin (Isabela)                             | 0.70278–0.70285                 | 0.51302–0.51304                   | 2        | 15*     |
| Wolf (Isabela)                               | 0.70269–0.70275                 | 0.51304–0.51308                   | 3        | 15*     |
| Fernandina                                   | 0.70312–0.70323                 | 0.51294                           | 4        | 15*     |
| Espanola                                     | 0.70294–0.70298                 | 0.51302–0.51305                   | 2        | 15*     |
| Santiago                                     | 0.70284–0.70294                 | 0.51298–0.51307                   | 4        | 15*     |
| Pinzon                                       | 0.70301–0.70311                 | 0.51302                           | 2        | 15*     |
| Rabida                                       | 0.70297–0.70302                 | 0.51301                           | 2        | 15*     |
| Floreana                                     | 0.70338–0.70366                 | 0.51291–0.51298                   | 5        | 13, 15* |
| Darwin (Island)                              | 0.70300–0.70329                 | 0.51295–0.51298                   | 2        | 15*     |
| Wolf (Island)                                | 0.70300–0.70337                 | 0.51295–0.51300                   | 2        | 15*     |
| Marchena                                     | 0.70276–0.70291                 | 0.51302–0.51307                   | 14       | 15, 18* |
| San Cristobal                                | 0.70282–0.70327                 | 0.51301–0.51309                   | 7        | 8, 15*  |
| Santa Fe                                     | 0.70293–0.70300                 | 0.51301–0.52303                   | 2        | 15*     |
| Roca Redonda                                 | 0.70311                         | 0.51296                           | 1        | *       |

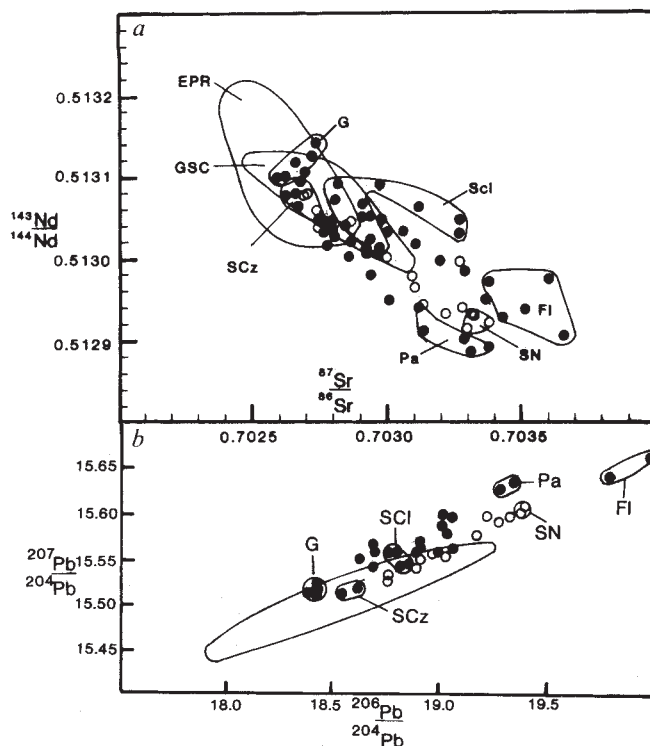
\* Also: W.M.W., manuscript in preparation.

trends. Second, volcanism is not strictly time-transgressive in the direction of plate motion. For example, half of San Cristobal Island was erupted roughly at the same time as Fernandina<sup>8</sup>, although these islands are the farthest from each other in the direction of plate motion. Furthermore, Santa Fe, which is the oldest island<sup>9</sup>, lies in the centre of the archipelago. Third, there is no consistent petrological and geochemical evolution of volcanism, as seems to be the case in Hawaii<sup>10,11</sup>.

Perhaps the most unusual feature of Galapagos Island lavas is the spatial distribution of petrological and geochemical lava types through the archipelago. A simple plume-asthenosphere mixing model might predict that the most 'hotspot-like' magmas would erupt in the centre of the archipelago (directly over the hotspot) and more mid-oceanic ridge basalt-like (MORB-like) basalts would erupt closer to the GSC as the plume mixed with the surrounding asthenosphere<sup>12</sup>. But in the Galapagos, basalts with strongly MORB-like petrologic features appear in the centre of the archipelago, and basalts with more enriched signatures appear around its margins (Fig. 1).

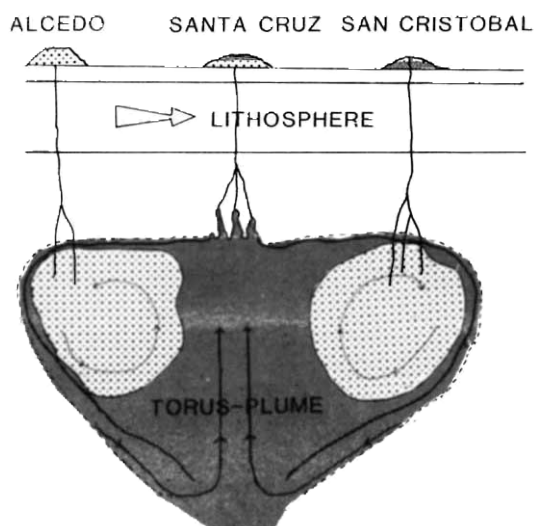
The geochemistry of Galapagos basalts ranges from tholeiites virtually indistinguishable from common MORBs to alkali-olivine basalts. Some of the volcanoes are made up of lavas with exceedingly small ranges in composition and others are strongly diverse. For example, Genovesa is composed entirely of MORB-like tholeiites and Floreana of incompatible trace-element enriched alkali-olivine basalts. San Cristobal, Santa Cruz, and Santiago each are made up of lavas that span nearly this entire compositional range, and the different types of lavas are randomly distributed on the younger parts of each island<sup>8,13,14</sup>.

Sr, Nd, and Pb isotopic ratios of lavas from the Galapagos Islands (ref. 15 and W.M.W., manuscript in preparation) (Table 1; Fig. 2) lie on an extension of the MORB trend<sup>16</sup>, but also have a broad overlap with the MORB fields, indicating time-averaged depletion in Rb relative to Sr and Nd relative to Sm. Most Galapagos lavas are not as depleted as common MORBs, but lavas from Genovesa and some lavas from Santa Cruz and Santiago Islands are indistinguishable from basalts erupted from the East Pacific Rise and GSC (Fig. 2). In detail, the isotope correlation diagrams display several interesting features. Three distinct subfields at the enriched end of the array are defined by the lavas of Pinta, Floreana, and Sierra Negra. Moreover, the geographic location of each volcano is related to its position



**Fig. 2** *a*, Sr-Nd isotopic variation among Galapagos lavas, also indicating lavas from the East Pacific Rise<sup>18</sup> (EPR) and Galapagos spreading centre<sup>7</sup> (GSC). Lavas from Isabela Island are indicated by open circles and the fields for several volcanoes indicated by letters (G = Genovesa; SCZ = Santa Cruz; SCI = San Cristobal; Pa = Pinta; SN = Sierra Negra; FI = Floreana). *b*, As above, but indicating Pb isotopic variation.

on these diagrams. The western volcanoes form a linear trend on the Sr-Nd isotope diagram extending through the centre of the array, with samples from Volcan Wolf at the depleted end and those from Sierra Negra at the enriched end. Similarly, Genovesa, Santa Cruz, Marchena, Darwin, Wolf and Pinta form a linear array<sup>17,18</sup>. The enriched source that contributes to each volcano therefore appears to be controlled by whichever of the three enriched-source volcanoes is closest geographically.



**Fig. 3** Schematic diagram of the toroidal diapir model of Griffiths<sup>3,4</sup> as applied to the Galapagos Islands, showing flow-lines. Dashed line indicates extend of thermal boundary layer. Stippled pattern indicates asthenospheric material entrained into the thermal.

White and Hofmann<sup>15</sup> originally observed a regional pattern of  $^{87}\text{Sr}/^{86}\text{Sr}$  in the Galapagos Archipelago, with  $^{87}\text{Sr}/^{86}\text{Sr} < 0.7030$  in the central islands and  $> 0.7030$  to the west, north, and south. This approach is extended herein with the newly acquired isotopic data and one major modification: instead of plotting the present location of the volcanoes, each measured lava is projected in the direction of plate motion by an amount determined by its age and the velocity of the Nazca Plate<sup>19</sup>. The locations of the lavas are therefore plotted relative to their absolute position when they erupted. When plotted in this way, both the high and low  $^{87}\text{Sr}/^{86}\text{Sr}$  ratios from each volcano (Table 1) form a consistent gradational pattern with only a few aberrant points (Fig. 1), with the centre of the archipelago being more depleted and the lavas becoming progressively more enriched outward.

Other workers<sup>7</sup> have argued that the isotopic variation within the Galapagos Islands is caused by mixing at shallow crustal levels, owing to the non-linearity of the sample points on a Sr-Nd isotope correlation diagram. Their reasoning was that the spread of points is explainable by mixing of two components with differing Sr/Nd elemental ratios but similar  $^{87}\text{Sr}/^{86}\text{Sr}$  and  $^{143}\text{Nd}/^{144}\text{Nd}$  isotopic ratios, as would be the case if variable plagioclase feldspar fractionation had taken place in the two mixing end members. Plagioclase is not on the liquidus in many of these lavas<sup>8,13</sup>, however, and the Pb isotopic data clearly indicate more than two end members. Non-linearity of the Sr-Nd isotopic array undoubtedly reflects non-binary mixing rather than variable Sr/Nd ratios, and mixing may have occurred in the mantle.

Griffiths<sup>3,4</sup> has determined how compositional and thermal contributions to buoyancy affect the flow pattern in ascending diapirs and finds that in diapirs (termed 'thermals') with ratios of compositional/thermal density contributions ( $\equiv R_p$ ) less than unity and large thermal Rayleigh numbers ( $> 10^2$ ), the diapirs entrain surrounding material into their centres during ascent. In a simplified sense, this is the result of the relatively high diffusivity of heat from the diapir. Heat from the diapir creates a boundary layer in the surrounding medium which decreases its density and viscosity, and it becomes rheologically and buoyantly like the diapir material. This heated material becomes entrained through the bottom of the thermal and is concentrated near its centre, while the original material moves towards the outside. Such a diapir would consist of incompatible-element enriched material surrounding an incompatible-element dep-

leted core (Fig. 3). An example of possible mantle material meeting these criteria is ancient subducted crust<sup>20</sup>, which would have a thermal Rayleigh number  $\sim 10^3$  and  $R_p > -1$  (ref. 4).

This model appears to explain many of the enigmatic features of the Galapagos. We suggest that a thermally buoyant (as opposed to compositionally buoyant diapir rising beneath the Galapagos entrains depleted asthenospheric material into its core as it traverses the upper mantle. The entrainment of asthenospheric material may be enhanced by the proximity of the GSC. Volcanoes in the centre of the Galapagos tap a core consisting of asthenosphere-derived material with small, but variable, admixtures of diapir material, whereas volcanoes on the periphery of the archipelago tap relatively undiluted diapir material. In addition, the enriched part of the diapir must be heterogeneous on a regional scale (as observed for the Hawaiian hotspot<sup>21</sup>), a feature necessary to explain the three enriched Galapagos end members.

The distribution of active volcanism throughout the Galapagos could occur if the diapir were roughly the diameter of the archipelago. It has been suggested<sup>22,6</sup> that contamination of the asthenosphere beneath the GSC results from diapir material being drawn to the ridge from the mantle beneath the Galapagos Islands through a sublithospheric channel. According to this hypothesis, the northern islands are a manifestation of this channel. One might expect that there would be progressive mixing along this channel, with the magmas becoming gradually more depleted towards the ridge. This is clearly not the case (Fig. 1). A diapir the size of the entire archipelago could mix directly with the asthenospheric source of GSC magmas, and no distant horizontal transport of diapir material would be required.

The temporal variation of San Cristobal and Santa Cruz Islands also supports the proposal of entrainment of depleted material into a toroidal thermal diapir, even though they have evolved in opposite directions. The oldest lavas from the Santa Cruz-Balra-North Seymour volcanic group are 2.2 to 0.6 Myr old (ref. 13) and have a relatively enriched isotopic signature ( $\epsilon_{\text{Nd}} = 7.3$  to 7.9). These lavas were erupted  $\sim 35$  to 120 km west of their present location (in the hotspot reference frame). The lavas constituting the younger shield of Santa Cruz have a more depleted isotopic signature ( $\epsilon_{\text{Nd}} = 8.6$ –9.0). The temporal evolution of Santa Cruz is therefore consistent with the hypothesis that the first magmas came from the sides of the diapir, which is dominated by the original diapir material, and the volcano passed over progressively more depleted material near the centre of the diapir. In contrast, the older shield of San Cristobal, which is 2.3–0.6 Myr old (ref. 8), is made up of relatively depleted lavas ( $\epsilon_{\text{Nd}} = 8.8$ ), and the younger lavas are relatively enriched ( $\epsilon_{\text{Nd}} = 8.3$ –7.1), consistent with the older series forming over the centre of the diapir and later lavas being derived from the more enriched outer part of the diapir. Thus the toroidal diapir model appears to explain contrasting temporal evolution of Galapagos volcanoes.

Whether this hypothesis of toroidal diapirs is applicable to other hotspots remains to be seen. Similar features have been reported from the Marquesas Island chain<sup>23</sup> and the Society Islands as well<sup>24</sup>.

This work was partially supported by NSF awards to W.M.W. and A.R.M.

Received 12 January; accepted 11 April 1988.

- Holden, J. C. & Dietz, R. S. *Nature* **235**, 266–269 (1972).
- Morgan, W. J. *Nature* **230**, 42–43 (1971).
- Griffiths, R. W. *Earth planet. Sci. Lett.* **78**, 435–446 (1986).
- Griffiths, R. W. *Phys. Earth planet. Inter.* **43**, 261–273 (1986).
- Schilling, J. G., Kingsley, R. H. & Devine, J. D. *J. geophys. Res.* **87**, 5593–5610 (1982).
- Verma, S. P. & Schilling, J. G. *J. geophys. Res.* **87**, 10838–10856 (1982).
- Verma, S. P., Schilling, J. G. & Wagoner, D. G. *Nature* **306**, 654–657 (1983).
- Geist, D. J., McBirney, A. R. & Duncan, R. A. *Geol. Soc. Am. Bull.* **97**, 555–566 (1986).
- Geist, D. J., McBirney, A. R. & Duncan, R. A. *J. Volcan. geotherm. Res.* **26**, 203–212 (1985).
- MacDonald, G. A. & Katsura, T. *J. Petrology* **5**, 82–133 (1964).
- Chen, C. Y. & Frey, F. A. *Nature* **302**, 785–789 (1983).
- Schilling, J. G. & Noe-Nygaard, A. *Earth planet. Sci. Lett.* **24**, 1–14 (1974).



13. Bow, C. S. thesis, Univ. Oregon (1979).
14. Swanson, F. J., Baitis, H. W., Lexa, J. & Dymond, J. *Geol. Soc. Am. Bull.* **85**, 1803–1810 (1974).
15. White, W. M. & Hofmann, A. W. *Carnegie Inst. Wash. Yearbook* **77** 596 (1978).
16. White, W. M. *Geology* **13**, 115–118 (1985).
17. Cullen, A. B. & McBirney, A. R. *Geol. Soc. Am. Bull.* **98**, 294–301 (1987).
18. Vicenzi, E. P., Cullen, A. B., McBirney, A. R. & White, W. M. *EOS* **66**, 1107 (1985).
19. Hey, R., Johnson, G. L. & Lowrie, A. *Geol. Soc. Am. Bull.* **88**, 1385–1403 (1977).
20. Hofmann, A. W. & White, W. M. *Earth planet. Sci. Lett.* **57**, 421–436 (1982).
21. Stille, P., Unruh, D. M. & Tatsumoto, M. *Geochim. cosmochim. Acta* **50**, 2303–2319 (1986).
22. Morgan, W. J. *J. geophys. Res.* **83**, 5355–5360 (1978).
23. McCulloch, M. T., Duncan, R. A. & Griffiths, R. W. *EOS* **66**, 1079 (1985).
24. White, W. M., Duncan, R. A., Geist, D. J. & McBirney, A. R. *EOS* **68**, 438 (1987).
25. Cox, A. & Dalrymple, G. B. *Nature* **209**, 776–777 (1966).
26. Bailey, K. *Science* **192**, 465–467 (1976).

## Mechanistic resource competition theory applied to laboratory experiments with zooplankton

Karl O. Rothhaupt

Max-Planck-Institut für Limnologie, Postfach 165, D-2320 Plön, FRG

There is a strong feeling that ecological thought today is hindered by the traditional, phenomenological Lotka–Volterra model. Recent publications have stressed the importance of considering the mechanisms underlying competitive interactions<sup>1,2</sup>. A mechanistic, resource-based approach to studying exploitative competition for limiting nutrients has been very successful in phytoplankton ecology<sup>3,4</sup>, where various nutrients are essential resources. Here I report results of laboratory experiments that show for the first time the applicability of this approach to the competition for substitutable resources in herbivorous zooplankton. The outcome of exploitative competition between the rotifers *Brachionus rubens* and *B. calyciflorus* for two species of food algae was predicted by a graphical model<sup>5</sup> in 11 of 12 cases. These findings extend the applicability of the mechanistic consumer-resource approach to animal populations, and illustrate the phenomenon of food-niche separation.

Ecological theory holds that as many consumer species can coexist in equilibrium communities as there are limiting resources<sup>6</sup>. This has been shown to be true for algae depending on non-substitutable resources<sup>7</sup>. To verify the premise for animal populations that compete for a common resource base, I performed laboratory experiments with the two rotifers *Brachionus*

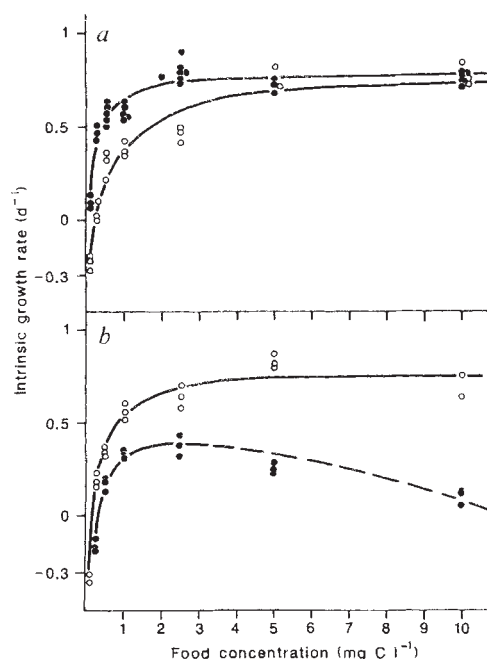


Fig. 1 Relationship between environmental food concentration and intrinsic growth rate for *B. rubens* (closed circles) and *B. calyciflorus* (open circles). *B. rubens* is the superior competitor (reaching any possible growth rate at a lower food concentration) for *Monoraphidium minutum* (a), whereas *B. calyciflorus* is always superior for *Chlamydomonas sphaeroides* (b). The decline in growth rate for *B. rubens* at high densities of *Chlamydomonas* is caused by mechanical disturbance of the feeding process. Algae and animals were grown in the same defined medium<sup>21</sup>. Food concentration was the independent variable, growth rate the dependent variable in these experiments. Standard food concentrations ( $\text{mg C l}^{-1}$ ) were 0.1 (except *B. rubens*–*Chlamydomonas*), 0.25, 0.5, 1, 2.5, 5 and 10. Stopped glass tubes holding 55 ml food suspension received 50 animals chosen at random from healthy stock cultures and were kept in the dark on a rotating wheel at 1 r.p.m. At the maximum, 25%, usually much less, of the volume was cleared per day. At 24-h intervals the food suspension was completely renewed and the rotifers were counted. Fifty animals of the experimental population (or the remaining ones in the case of negative growth rates) were pipetted back into the tube. Care was taken not to change the established age distribution. When per cent increase (decrease) had stabilized for at least three days (coefficient of variance  $\leq 0.1$ ), the experiment was terminated. Instantaneous growth rates ( $r, \text{d}^{-1}$ ) were calculated as  $r = \ln(N_t) - \ln(N_{t-1})$ . The curves are Monod curves<sup>22</sup> with a threshold for zero-growth, fitted by nonlinear regression. In the case of *B. rubens* growing on *Chlamydomonas*, the curve is fitted only to the data below the growth-rate maximum.

Table 1 Clearance rates and feeding selectivities

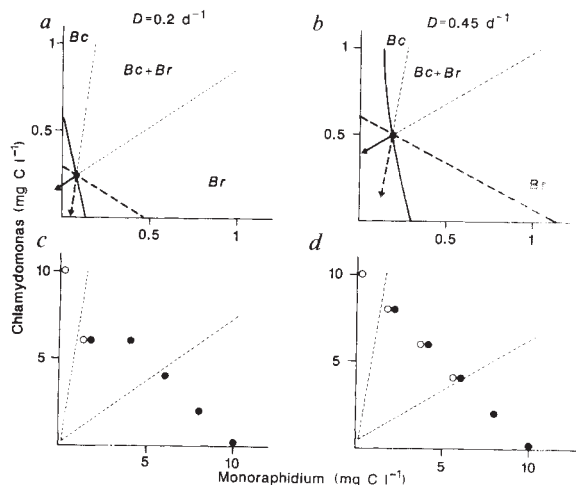
|                        | Maximal clearance rate on: |                       |                        |
|------------------------|----------------------------|-----------------------|------------------------|
|                        | <i>Monoraphidium</i>       | <i>Chlamydomonas</i>  | $D_{\text{Mono}}$      |
| <i>B. rubens</i>       | $10.50 \pm 0.31$ (38)      | $2.04 \pm 0.21$ (15)  | $0.658 \pm 0.030$ (4)  |
| <i>B. calyciflorus</i> | $8.78 \pm 0.36$ (20)       | $17.78 \pm 1.32$ (11) | $-0.526 \pm 0.026$ (3) |

Numbers are means  $\pm 1$  s.e.m. (number of observations in parentheses). All experiments were performed at  $20 \pm 1^\circ \text{C}$ . Maximal clearance rates ( $\mu\text{l}$  per animal  $\text{h}^{-1}$ ) are averages of measurements below a critical food concentration (ILL), up to which ingestion rates increase linearly with food concentration, indicating constant clearance rates. The ILL changes with rotifer species and food algae (unpublished results). Clearance rates were determined in short-term feeding experiments with  $^{14}\text{C}$ -labelled algae, at the same food concentrations as the growth rates (Fig. 1). About 100 animals from healthy stock cultures were pre-fed with unlabelled algae at the respective concentrations for at least 1 h. Exposure times to radioactive food were 10 min in 20 ml experimental volume. Then the animals were drugged in carbonated water, fixed in formaldehyde and groups of 12 animals were immediately transferred into scintillation vials. Scintillation counting and calculation of clearance and ingestion rates followed standard procedures<sup>18</sup>.

Selectivities in mixed food were expressed as  $D$  (ref. 19).  $D$  ranges from  $-1$  to  $+1$ .  $D_{\text{Mono}} = 0$  indicates unselective feeding, positive values indicate selection for, negative values indicate selection against *Monoraphidium*. Feeding in these experiments was determined by the closed-bottle method<sup>18</sup> with 25 ml experimental volume, three control bottles without animals, 125 animals per test vessel, 24-h exposure time,  $0.4 \text{ mg C l}^{-1}$  *Chlamydomonas*,  $0.2 \text{ mg C l}^{-1}$  *Monoraphidium*. Algae were counted under an inverted microscope<sup>20</sup> (at least  $2 \times 400$  cells per sample for *Monoraphidium* or two diameters of the counting chamber for *Chlamydomonas*).

*rubens* and *B. calyciflorus*. These two species differ but overlap in their preferred food size spectra. When either *Monoraphidium minutum* (Chlorophyta, equivalent spherical diameter,  $3.5 \mu\text{m}$ ) or *Chlamydomonas sphaeroides* (Chlorophyta, equivalent spherical diameter,  $12 \mu\text{m}$ ) are offered, *B. rubens* has higher maximal clearance rates on the smaller alga, whereas *B. calyciflorus* feeds more efficiently on the bigger one (Table 1). When both food algae are offered simultaneously, each rotifer species shows selectivity in favour of the respective better ingestible particle (Table 1), reflecting different retention efficiencies for particles of different sizes (my unpublished data). The relative ingestibilities of both food algae influence their suitabilities to maintain population growth. When grown on one alga, the rotifer species that more efficiently ingests that food item is superior (Fig. 1).

With these numerical responses (population growth rates) as well as the functional responses (clearance rates) to different types and densities of food resource competition theory predicts

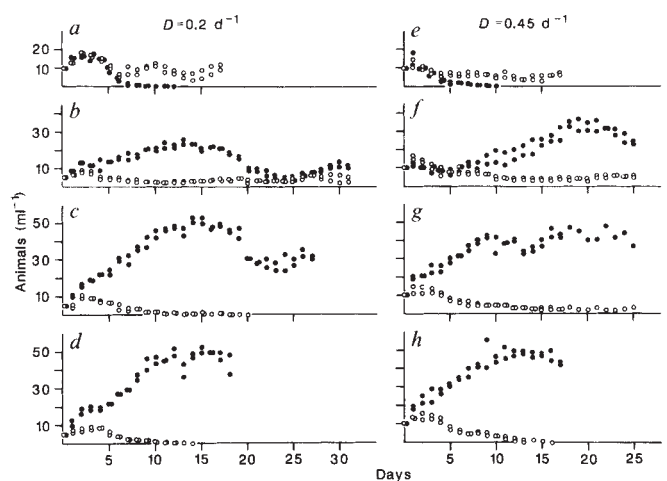


**Fig. 2** Theoretical predictions of the outcome of exploitative competition between *B. rubens* (*Br*) and *B. calyciflorus* (*Bc*) at two different dilution rates ( $D$ ). In the upper panel (*a, b*), the zero net growth isocline and the consumption vector are drawn in solid lines for *B. rubens* and in broken lines for *B. calyciflorus*. Both species are expected to coexist ( $Bc + Br$ ) only for resource concentrations within the sector between the thin hatched lines. In the lower panel (*c, d*; different scale) these predictions are compared to the actual outcome of competition. Solid dots, experiments in which *B. rubens* persisted; open circles, persistence of *B. calyciflorus*; both symbols side-by-side, coexistence. Theory accurately predicts the outcome of competition in 11 of 12 cases. With food concentrations of 6:4 (*Chlamydomonas*  $\text{mgC l}^{-1}$ : *Monoraphidium*  $\text{mgC l}^{-1}$ ) at the lower dilution rate ( $D = 0.2 \text{ d}^{-1}$ , *2c*) *B. rubens* took over, whereas coexistence was predicted.

the outcome of competition (Fig. 2*a, b*). At equilibrium, the animals' intrinsic growth rate ( $r$ , per day ( $\text{d}^{-1}$ )) equals the dilution rate ( $D$ , relative replacement of culture volume per unit time,  $\text{d}^{-1}$ ). Following the graphical method developed by Tilman<sup>5</sup>, growth isoclines are plotted on a resource plane. The zero net growth isoclines (ZNGIs) denote the resource concentrations at which reproduction exactly balances the losses ( $r = D$ ). In natural situations losses can be given by predation, parasitism or emigration. If two species' ZNGIs do not intersect, the species that has the ZNGI closer to the origin will be the superior competitor. If ZNGIs cross, it is important to consider both species' ratios of resource consumption, which can be represented as consumption vectors. The intersection point marks a possible stable equilibrium if at that point each species consumes relatively more of the resource that more limits its own growth. The species will coexist if the resource supply concentrations fall within a sector of the resource plane that is delimited by the slopes of the consumption vectors at the equilibrium point. Only then will the availabilities of both resources be focused into the two-species equilibrium point by the combined effects of resource consumption and resource supply<sup>5</sup>.

In the case I report, assuming perfectly substitutable resources, the ZNGIs are constructed by connecting the zero net growth points on the resource axes with a straight line. In the case of  $D = 0.45 \text{ d}^{-1}$ , no zero net growth point exists for *B. rubens* on the *Chlamydomonas* resource axis. Here the equilibrium point is constructed using the condition that the sum of ingestion of both food algae must be sufficient to maintain the necessary growth rate. The consumption vectors are calculated as the products of clearance rates (Table 1) and estimated equilibrium algal densities.

Examples for the population dynamics are shown in Fig. 3. The comparison of prediction and experimental outcome shows excellent agreement for the higher dilution rate ( $D = 0.45 \text{ d}^{-1}$ , Fig. 2*d*). Theory was successful in predicting both coexistence and exclusion correctly. At the lower dilution rate ( $D = 0.2 \text{ d}^{-1}$ ,



**Fig. 3** Examples for the time course of competition experiments at two different dilution rates ( $D$ ). Solid dots denote *B. rubens*, open circles denote *B. calyciflorus*. *B. rubens* wins in *c, d* and *h*. *B. calyciflorus* wins in *a* and *e*. Coexistence occurs in *b, f* and *g*. *B. calyciflorus* always establishes lower population numbers than *B. rubens*. This is due to a higher biomass and a lower (carbon-related) growth efficiency of *B. calyciflorus*.

**Methods.** Competition experiments were run in duplicate with the same basic setup as explained for Fig. 1, except that every day only a certain amount of medium (10 or 20 ml, including animals) was removed and replaced with fresh food suspension. Equilibrium growth rates were determined by the average dilution rate resulting from daily collection<sup>23</sup>. Rotifer populations were allowed to increase and exploit the food resources. The criterion for coexistence was, that for a timespan of at least three turnover-times ( $T = D^{-1}$ ), the slope of the regression of population numbers versus days for the rarer species was not significantly different from zero. Resource supply concentrations (*Chlamydomonas*  $\text{mgC l}^{-1}$ : *Monoraphidium*  $\text{mgC l}^{-1}$ ) were:  $D = 0.2 \text{ d}^{-1}$ , 10:0 (*a*), 6:1.5(*b*), 4:6(*c*), 2:8(*d*);  $D = 0.45 \text{ d}^{-1}$ , 10:0(*e*), 8:2(*f*), 4:6(*g*), 2:8(*h*).

Fig. 2*c*), in five of six cases prediction and experimental outcome correspond. Here the region of coexistence is expected to be narrower, which can be seen by comparing the predictions for the supply points at 6:4 (*Monoraphidium*  $\text{mgC l}^{-1}$ : *Chlamydomonas*  $\text{mgC l}^{-1}$ , Fig. 2*c, d*). The experimental outcome followed that trend, but with supply concentrations of 4:6 *B. rubens* took over although coexistence was expected.

Regarding the very low equilibrium densities of *B. calyciflorus*, the lack of agreement in this one case could have been caused by stochastic extinction. But the experimental outcome occurred in two independent replicates and the experimental volume (55 ml) seems to assure a sufficiently high population number even at comparatively low population densities of only a few animals  $\text{ml}^{-1}$ . Another explanation could be that the assumption of perfectly substitutable resources is not met. If both food algae are substitutable but complementary this would change the shape of the ZNGIs<sup>5</sup> and shift the location of the equilibrium point on the resource plane. As the slopes of the consumption vectors are influenced by the resource availabilities, this would also influence the boundaries of predicted experimental outcomes.

This is the first experimental demonstration of the applicability of Tilman's resource-based competition theory<sup>5</sup> to animal populations, an approach that may help elucidate mechanisms of niche separation. Despite the wide acceptance of the niche concept<sup>8</sup>, there are still few direct examples for niche-based coexistence in animals<sup>9-15</sup>. The main obstacle has been the difficulty of reconciling the mechanisms of competition with the still pervasive Lotka-Volterra model<sup>1</sup>. Taking the mechanisms of consumer-resource interactions explicitly into consideration



may generally prove to be a powerful approach. This is certainly the case in zooplankton ecology. Research in this area has been largely predation-oriented and better understanding of competitive interactions is needed<sup>16</sup>. Many other influences may lead to complications *in situ*, for example the presence of noxious or interfering algae, intricate responses of algal communities to zooplankton grazing<sup>17</sup>, or nonequilibrium conditions. However, according to our present knowledge of zooplankton feeding strategies, this mechanistic approach, so successful in phytoplankton ecology, also looks promising for predicting taxonomic trends in zooplankton occurrence.

I thank W. Lampert, for advice and encouragement, U. Sommer from whom I learned about resource competition theory, B. Santer for help with the experiments and W. Lampert, R. W. Sterner, S. S. Kilham and P. Kilham for comments on the manuscript.

Received 24 December 1987; accepted 21 April 1988.

1. Tilman, D. *Am. Nat.* **129**, 769–774 (1987).
2. Holt, R. D. *Oikos* **48**, 110–114 (1987).
3. Tilman, D., Kilham, S. S. & Kilham, P. A. *Rev. Ecol. Syst.* **13**, 349–372 (1982).
4. Sommer, U. *Limnol. Oceanogr.* **30**, 335–346 (1985).
5. Tilman, D. *Resource Competition and Community Structure* (Princeton University Press, 1982).
6. Armstrong, R. A. & McGehee, R. *Am. Nat.* **115**, 151–170 (1980).
7. Tilman, D. *Ecology* **58**, 338–348 (1977).
8. Schoener, T. W. in *Community Ecology: Pattern and Process* (eds Kikkawa, J. & Anderson, D. J.) 91–126 (Blackwell Scientific, Oxford, 1986).
9. Reynoldson, T. B. & Bellamy, L. S. *Oikos* **24**, 301–313 (1973).
10. Reynoldson, T. B., Gilliam, J. F. & Jaques, R. M. *Arch. Hydrobiol.* **92**, 71–113 (1981).
11. Werner, E. E. & Hall, D. J. *Science* **191**, 404–406 (1976).
12. Werner, E. E. & Hall, D. J. *Ecology* **58**, 869–876 (1977).
13. Werner, E. E. & Hall, D. J. *Ecology* **60**, 256–264 (1979).
14. DeMott, W. R. & Kerfoot, W. C. *Ecology* **63**, 1949–1966 (1982).
15. Arthur, W. & Middlecote, J. *Biol. J. Linn. Soc.* **23**, 167–176 (1984).
16. Lampert, W. *Arch. Hydrobiol./Beihft* **21**, V–VII (1985).
17. Sterner, R. W. *Science* **231**, 605–607 (1986).
18. Peters, R. H. in *A Manual on Methods for the Assessment of Secondary Productivity in Fresh Waters*, IBP Handbook 17 Edn 2 (eds Downing J. A., & Rigler, F. H.) 336–412 (Blackwell Scientific, Oxford, 1984).
19. Jacobs, J. *Oecologia* **14**, 413–417 (1974).
20. Utermöhl, H. *Mitt. Int. Verein. theor. angew. Limnol.* **9**, 1–39 (1958).
21. Müller, H. *Arch. Hydrobiol. Suppl.* **38**, 399–484 (1972).
22. Monod, J. *Ann. Inst. Pasteur Lille* **79**, 390–410 (1950).
23. Gaedeke, A. & Sommer, U. *Oecologia* **71**, 25–28 (1986).

## Volatile general anaesthetics activate a novel neuronal K<sup>+</sup> current

N. P. Franks & W. R. Lieb

Biophysics Section, Blackett Laboratory, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BZ, UK

Although it is still controversial whether the primary target sites underlying general anaesthesia are proteins<sup>1,2</sup> or lipids<sup>3–5</sup>, it is generally thought that the ultimate targets are ion channels in nerve membranes. One approach to finding these targets is to study the effects of general anaesthetics on identified neurons, where differential effects on neuronal activity can be pursued to the molecular level. Here we report that amongst a group of apparently identical molluscan neurons having endogenous firing activity, a single cell displays an unusual sensitivity to volatile agents (which, at surgical levels, completely inhibit its activity). We further show that this sensitivity is due to a novel anaesthetic-activated K<sup>+</sup> current, which is found in the sensitive cell but not in the surrounding insensitive cells. This K<sup>+</sup> conductance is not appreciably voltage-gated and persists for as long as the anaesthetic is present. The response to anaesthetics is completely reversible and saturates at low anaesthetic partial pressures: the half-maximal response for halothane occurs at 0.0063 atm, close to its minimum alveolar concentration (0.0075 atm) in man<sup>6</sup>.

We have been studying an identified cluster of spontaneously active neurons<sup>7</sup> located in a lobe-like protrusion in the ventral region of the right parietal ganglion of the great pond snail

*Lymnaea stagnalis*. This cluster contains 30–50 neurons, which range in diameter from 20 to 100  $\mu\text{m}$  and exhibit a pattern of spontaneous firing which has been shown<sup>8</sup> to be endogenous in origin. We found that almost all of these neurons show little alteration in their steady-state firing activity when exposed to surgical ED<sub>50</sub> partial pressures (0.0080 atm) of the widely used volatile general anaesthetic halothane (CF<sub>3</sub>CHBrCl).

Remarkably, however, one of these neurons is extremely sensitive to anaesthetic inhibition (see Fig. 1a). This cell is one of the five largest neurons in the cluster, ~80  $\mu\text{m}$  in diameter and usually located adjacent to the largest cell. In the presence of surgical partial pressures of the volatile agents halothane (0.0080 atm), isoflurane (0.010 atm), chloroform (0.0066 atm) and ether (0.019 atm), the cell membrane potential hyperpolarized far below ( $\Delta V_{\text{mem}} \approx -25$  mV) the threshold potential needed to initiate firing of action potentials. Normal firing activity resumed on removal of the anaesthetic. Thus the effect is both large and reversible at surgical levels. Interestingly, the responses were very much more rapid when the cell body was slightly separated from the rest of the ganglion, exposing the base of the cell directly to the flowing solution. This suggests that the anaesthetic-binding sites are located near the axon hillock.

The anaesthetic-induced hyperpolarization persisted when external Cl<sup>-</sup> was replaced with isethionate, sulphate and propionate. It also remained when Na<sup>+</sup> was replaced with TRIS or tetraethylammonium (TEA<sup>+</sup>) or when Ca<sup>2+</sup> was replaced with Co<sup>2+</sup>. (The hyperpolarization was also unaffected by the presence of 100  $\mu\text{M}$  ouabain.) It thus appeared that anaesthetics were activating an outward K<sup>+</sup> current. Perhaps the best characterized<sup>9,10</sup> outward K<sup>+</sup> currents in neurons are the delayed rectifier current  $I_K$ , the fast transient potassium current  $I_A$ , and the calcium-activated potassium currents  $I_{K(\text{Ca})}$ . All of these currents were readily identified in our neurons but were found to be almost completely blocked in Na<sup>+</sup>-free and Ca<sup>2+</sup>-free solutions containing 50 mM TEA<sup>+</sup>, 3 mM 4-aminopyridine (4AP), and 8 mM Co<sup>2+</sup>. Under these conditions, nonetheless, anaesthetics still activated a large outward potassium current (see Fig. 1b) which was found to be relatively insensitive to these concentrations of TEA<sup>+</sup>, 4AP and Co<sup>2+</sup>. It can be seen that, at the normal K<sup>+</sup> concentration (2.5 mM), halothane activates an outward current which reverses at about -80 mV, close to the expected equilibrium potential for K<sup>+</sup> ions. Figure 1b shows that this current is indeed sensitive to the external K<sup>+</sup> concentration: when this was increased fourfold (to 10 mM K<sup>+</sup>), the reversal potential shifted 28 mV in the depolarizing direction (close to, but somewhat less than, the 35 mV predicted by the Nernst equation).

We found a direct correlation between the presence of this anaesthetic-induced current (which we call  $I_{K(\text{An})}$ ) and anaesthetic inhibition of spontaneous firing in a given neuron. Those neurons which were insensitive to anaesthetic lacked this  $I_{K(\text{An})}$  current, while it was always present in the sensitive cell. (We occasionally found neurons which displayed only a modest sensitivity to anaesthetic, and these neurons were found to have correspondingly small  $I_{K(\text{An})}$  currents.) Concentrations of anaesthetic which completely blocked spontaneous firing in the sensitive neuron had little effect on the shape of the action potentials in the insensitive neurons. Furthermore, injection of inward current into the anaesthetic-inhibited sensitive cell could restore neuronal firing. Thus it is the anaesthetic-induced hyperpolarization itself which is responsible for the inhibition of neuronal firing in the sensitive cell.

The voltage dependence of  $I_{K(\text{An})}$  is shown in Fig. 2. The solid line gives the current predicted for a K<sup>+</sup>-selective channel obeying the Goldman-Hodgkin-Katz constant-field equation; comparison with the experimental data shows that the anaesthetic-activated K<sup>+</sup> conductance can be largely accounted for by constant-field considerations and suggests that it is only slightly (if at all) voltage-gated. This interpretation is supported by the results of voltage-jump experiments. We find that  $I_{K(\text{An})}$  was

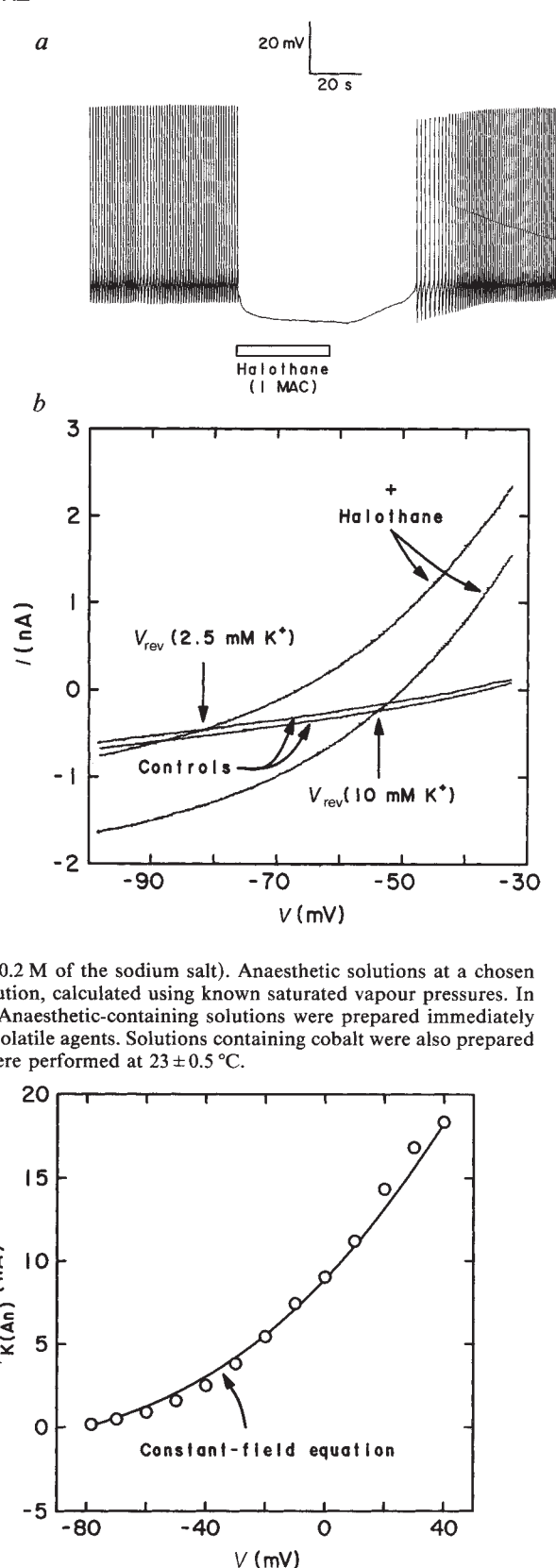
**Fig. 1** Surgical levels of the volatile general anaesthetic halothane reversibly block the endogenous firing activity of an identified neuron by activating an outward current. *a*, Continuous intracellular recording of membrane potential as a function of time before, during (bar) and after exposure of the sensitive cell to normal saline solution containing halothane at a partial pressure of 0.0080 atm (~minimal alveolar concentration). *b*, Steady-state  $I-V$  curves for the sensitive cell at two different  $K^+$  concentrations in the presence and absence of 0.010 atm halothane. Measurements were made in TEA/cobalt saline containing 2.5 mM and 10 mM  $K^+$ .

**Methods.** *Lymnaea stagnalis* (between 24 mm and 32 mm in length) were obtained from Blades Biological Ltd. (Cowden, Kent). The central nervous system was excised and the loose connective tissue overlying the right parietal ganglion removed. This ganglion was then dissected away intact and held by its nerve roots using light suction. The tough endoneurium was removed after softening in pronase (2 mg ml<sup>-1</sup>) for 10–15 min. This exposed the readily identifiable cluster of neurons in the ventral part of the ganglion used in these experiments. These cells are light yellow in colour and are thought to be neurosecretory<sup>7</sup>. The ganglion was held in a small chamber (volume 50  $\mu$ l) and continuously perfused (at 1–2 ml min<sup>-1</sup>). Solutions could be exchanged within seconds, but 2–4 min were usually allowed for complete equilibration. Electrodes, pulled using 1 mm and 1.5 mm glass capillaries, were filled with 2.5 M KCl and had resistances between 1.5 and 25 M $\Omega$ . A two-electrode voltage clamp circuit was employed, with the current record being filtered by an 8-pole Bessel filter. Data were stored on tape using a video recorder (Sony SL-F30) and a modified digital audio processor (Sony PCM 701ES). Data were subsequently digitized at 1.4 kHz and transferred to a computer for analysis. Steady-state  $I-V$  curves were constructed either by measuring the clamp current (filtered at 1 kHz corner frequency) 10–15 seconds after a voltage jump from a hyperpolarized membrane potential (usually -80 mV) to the various test potentials or by slowly ramping the membrane potential (usually at 3 mV s<sup>-1</sup>) in the depolarizing direction (with current records filtered at 100 Hz or 10 Hz corner frequency). Ramps performed either slower (1 mV s<sup>-1</sup>) or faster (5 mV s<sup>-1</sup>) gave essentially identical results. Intracellular ionophoretic injections were performed under voltage-clamp using a third electrode connected to a current pump. The composition of the normal saline was: NaCl, 50 mM; KCl, 2.5 mM; CaCl<sub>2</sub>, 4 mM; MgCl<sub>2</sub>, 4 mM; HEPES, 10 mM; glucose, 5 mM. In the TEA/cobalt saline, NaCl and CaCl<sub>2</sub> were replaced by 50 mM TEA-Cl (tetraethylammonium chloride) and 8 mM CoCl<sub>2</sub>. The TEA/4AP/cobalt saline had an additional 3 mM 4AP (4-aminopyridine). The normal saline was titrated to pH 7.4 with NaOH and the TEA/cobalt and TEA/4AP/cobalt saline solutions titrated to pH 7.4 with TEA-OH. For intracellular ionophoretic injection, the third electrode was filled with the following solutions: EGTA (0.5 M, titrated to pH 7.4 with KOH), TEA-Cl (0.5 M), BaCl<sub>2</sub> (1.0 M) or cyclic AMP (0.2 M of the sodium salt). Anaesthetic solutions at a chosen partial pressure were prepared using the appropriate fraction of a saturated solution, calculated using known saturated vapour pressures. In the case of diethyl ether, neat anaesthetic was added directly to the solutions. Anaesthetic-containing solutions were prepared immediately before the experiments and care was taken throughout to minimize losses of the volatile agents. Solutions containing cobalt were also prepared just before the experiments. All experiments were performed at 23  $\pm$  0.5  $^{\circ}$ C.

close to its steady-state value at the earliest times (2–5 ms) after the voltage jump at which we could reliably make measurements. Not only do we find no measurable delay in activating  $I_{K(An)}$ , but also little subsequent inactivation with time (see Fig. 3b).

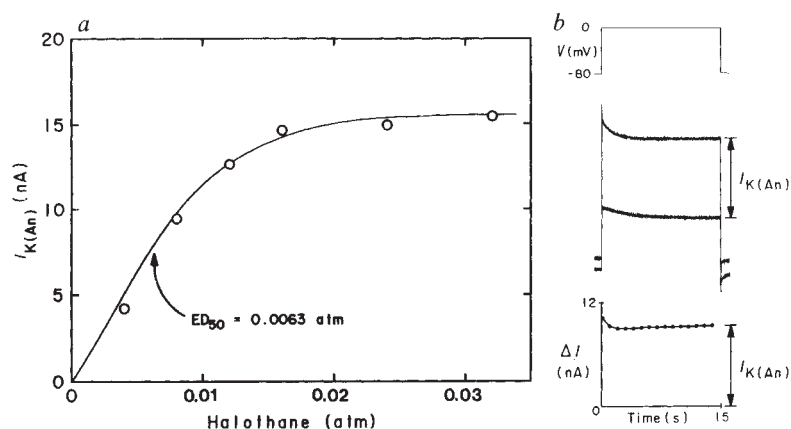
The dose-response curve of  $I_{K(An)}$  for halothane is shown in Fig. 3a. Interestingly, the response saturates at low levels of anaesthetic. This behaviour is consistent with, but certainly does not prove, the idea that anaesthetics are binding to saturable sites on protein molecules<sup>11</sup>. The half-maximal effect occurs at only 0.0063 atm halothane, which is comparable to the surgical ED<sub>50</sub> value<sup>6</sup> of 0.0075 atm for man and the anaesthetic ED<sub>50</sub> concentration for *Lymnaea stagnalis* (0.0083 atm) obtained<sup>12</sup> using the withdrawal reflex as the anaesthetic endpoint.

To further characterize  $I_{K(An)}$ , we used a third electrode to ionophoretically inject EGTA, TEA<sup>+</sup> and Ba<sup>2+</sup> into the sensitive cell, using the voltage-clamp circuit as a current sink. We found that  $I_{K(An)}$  was unchanged at internal levels of the Ca<sup>2+</sup>-chelating agent EGTA (50 nA, 4 min) which substantially blocked  $I_{K(Ca)}$ . The anaesthetic-induced current,  $I_{K(An)}$ , however, was reduced by ~50% (measured at  $V_{mem} = 0$  mV) by both internal TEA<sup>+</sup> (50 nA, 3 min) and internal Ba<sup>2+</sup> (2 nA, 4 min). In these and certain other respects (ionic dependence, insensitivity to the external blockers TEA<sup>+</sup> and 4AP, only slight voltage-gating, and no inactivation at depolarized potentials),  $I_{K(An)}$  resembles a serotonin-sensitive K<sup>+</sup> current ( $I_S$ ) found in both *Aplysia* and *Helix* neurons<sup>13–16</sup>. Bath application of serotonin (25  $\mu$ M) or intracellular injection of cyclic AMP (80 nA, 4 min), however,



**Fig. 2** The voltage dependence of the anaesthetic-activated current,  $I_{K(An)}$ . The current ( $\circ$ ) activated by halothane (0.0080 atm) as a function of voltage agrees reasonably well with that predicted (solid line) by the Goldman-Hodgkin-Katz constant-field equation<sup>9</sup> for a non-voltage-gated K<sup>+</sup> channel. The equation was evaluated assuming an internal K<sup>+</sup> concentration of 58 mM and a value for the product of the membrane permeability coefficient multiplied by the area of  $1.66 \times 10^{-9}$  cm<sup>3</sup> s<sup>-1</sup>. Measurements were made in TEA/4AP/cobalt saline.





**Fig. 3** *a*, The halothane dose-response curve for the anaesthetic-activated current,  $I_{K(An)}$ . Data were obtained in TEA/4AP/cobalt saline by measuring the current 15 s after a voltage jump from a holding potential of  $-80$  mV to a membrane potential of  $0$  mV. This membrane potential was chosen to minimize errors due to changes in non-specific leakage currents, which would be expected to reverse at  $\sim 0$  mV. To minimize systematic errors, the data points were obtained in random order. *b*, Voltage-jump records for membrane potential  $V$  and current  $I$  obtained as described above in the presence and absence of  $0.0080$  atm halothane. The anaesthetic-activated current ( $\Delta I$ ) is shown at the bottom.  $I_{K(An)}$  was defined as shown.

both of which are known<sup>13,15</sup> to close S channels, had little effect on  $I_{K(An)}$ . Furthermore, bath application of a saturated solution of arachidonic acid, which reportedly<sup>17</sup> opens S channels, had no effect on the sensitive cell. Thus the anaesthetic-activated current  $I_{K(An)}$  does not appear to share the same regulatory pathways as  $I_S$ , although there could be close similarities in the actual channel proteins. Single channel recording will help to clarify this issue.

The discovery of a large and persistent  $K^+$  current, activated at surgical  $ED_{50}$  levels of volatile general anaesthetics and present in some neurons but not others, has important implications for proposed mechanisms of action of these agents. The fact that anaesthetics activate this current in some neurons but not in others, and that this single current can completely inhibit neuronal firing, shows that these agents can act far more selectively than previously thought, at both the cellular and molecular level. The idea that volatile general anaesthetics might act by enhancing potassium currents is not a new one<sup>18,19</sup>, but until now the currents involved have not been characterized. (The hypothesis<sup>20</sup> that anaesthetics release calcium from internal stores and thus cause an increase in  $I_{K(Ca)}$  is not supported by our results.) The activation of a persistent  $K^+$  conductance could result in a generalized decrease in neuronal excitability, at both the pre- and post-synaptic levels, leading to a state of general anaesthesia.

How, at the molecular level, do anaesthetics activate  $I_{K(An)}$ ? It could be that anaesthetic molecules bind directly to the relevant channel and stabilize it in its open state. It is equally likely, however, that the channel is regulated by a second messenger system and anaesthetics act at some point in this pathway. Future studies should soon resolve this question. The presence of currents similar to  $I_{K(An)}$  in the central nervous systems of higher animals remains, of course, to be demonstrated. It is clear, however, that such anaesthetic-activated potassium channels have just the properties that might be expected for the principal target sites involved in general anaesthesia.

We thank Professor David Blow, Professor David Colquhoun, Dr Stuart Cull-Candy and Professor Wilfred Stein for helpful comments on the manuscript, Ted Dyett and Ian Coole for technical advice and Professor Eugene Dowdle and particularly Dr Stuart Cull-Candy for invaluable guidance. We are grateful to the MRC and the BOC Group, Inc. for support.

Received 23 March; accepted 12 May 1988.

- Richards, C. D. in *Topical Reviews in Anaesthesia* (eds Norman, J. & Whitwam, J. G.) Vol. 1, 1-84 (Wright, Bristol, 1980).
- Franks, N. P. & Lieb, W. R. *Nature* **300**, 487-493 (1982).
- Dluzewski, A. R., Halsey, M. J. & Simmonds, A. C. *Molec. Aspects Med.* **6**, 459-573 (1983).
- Haydon, D. A., Elliott, J. R. & Hendry, B. M. *Curr. Top. Membranes Transp.* **22**, 445-482 (1984).
- Miller, K. W. *Int. Rev. Neurobiol.* **27**, 1-61 (1985).
- Steward, A., Allott, P. R., Cowles, A. L. & Mapleson, W. W. *Br. J. Anaesth.* **45**, 282-293 (1973).
- Wendelaar Bonga, S. E. Z. *Zellforsch. mikrosk. Anat.* **108**, 190-224 (1970).
- van Swigchem, H. J. *exp. Biol.* **80**, 55-67 (1979).

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1984).
- Adams, W. B. & Benson, J. A. *Prog. Biophys. molec. Biol.* **46**, 1-49 (1985).
- Franks, N. P. & Lieb, W. R. *Nature* **310**, 599-601 (1984).
- Cruickshank, S. G. H., Girdlestone, D. & Winlow, W. *J. Physiol., Lond.* **367**, 8P (1985).
- Deterre, P., Paupardin-Tritsch, D., Bockaert, J. & Gerschenfeld, H. M. *Nature* **290**, 783-785 (1981).
- Klein, M., Camardo, J. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5713-5717 (1982).
- Siegelbaum, S. A., Camardo, J. S. & Kandel, E. R. *Nature* **299**, 413-417 (1982).
- Shuster, M. J. & Siegelbaum, S. A. *J. gen. Physiol.* **90**, 587-608 (1987).
- Piomelli, D. et al. *Nature* **328**, 38-43 (1987).
- Krnjević, K. *Anesthesiology* **34**, 215-217 (1971).
- Nicoll, R. A. & Madison, D. V. *Science* **217**, 1055-1057 (1982).
- Krnjević, K. in *Cellular Biology and Toxicity of Anesthetics* (ed. Fink, B. R.) 3-9 (William & Wilkins, Baltimore, 1972).

## Vascular endothelial cells synthesize nitric oxide from L-arginine

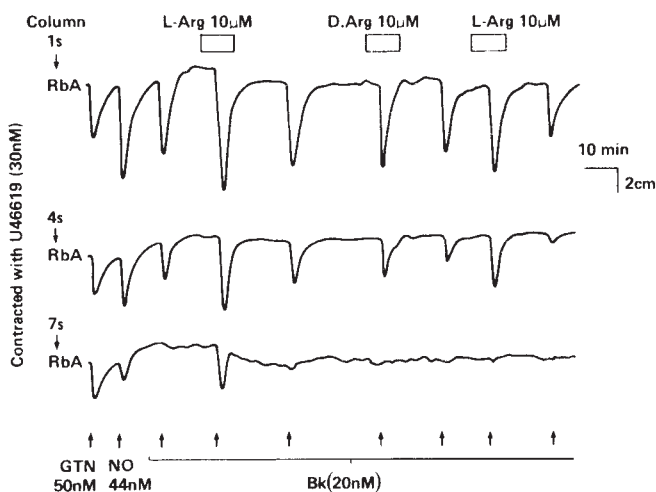
R. M. J. Palmer, D. S. Ashton & S. Moncada\*

Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

Nitric oxide (NO) released by vascular endothelial cells accounts for the relaxation of strips of vascular tissue<sup>1</sup> and for the inhibition of platelet aggregation<sup>2</sup> and platelet adhesion<sup>3</sup> attributed to endothelium-derived relaxing factor<sup>4</sup>. We now demonstrate that NO can be synthesized from L-arginine by porcine aortic endothelial cells in culture. Nitric oxide was detected by bioassay<sup>5</sup>, chemiluminescence<sup>1</sup> or by mass spectrometry. Release of NO from the endothelial cells induced by bradykinin and the calcium ionophore A23187 was reversibly enhanced by infusions of L-arginine and L-citrulline, but not D-arginine or other close structural analogues. Mass spectrometry studies using <sup>15</sup>N-labelled L-arginine indicated that this enhancement was due to the formation of NO from the terminal guanidino nitrogen atom(s) of L-arginine. The strict substrate specificity of this reaction suggests that L-arginine is the precursor for NO synthesis in vascular endothelial cells.

Porcine aortic endothelial cells were cultured on microcarriers for 10-16 days<sup>5</sup>. The cells were then washed three times and cultured for a further 24 h in fresh culture medium with or without L-arginine. These cells are designated CM cells or CM(-Arg) cells respectively. The release of NO was detected by bioassay on a series of rabbit aortic strips<sup>5</sup>, by chemiluminescence<sup>1</sup>, or as <sup>15</sup>NO by mass spectrometry. The chemiluminescence method detects both NO and NO<sub>2</sub><sup>-</sup>; identification of the released material as NO is based on the correspondence between this method and the bioassay which only detects NO<sup>1</sup>. Prostacyclin was measured by radioimmunoassay of its stable breakdown product, 6-keto-PGF<sub>1α</sub> (ref. 6).

\* To whom correspondence should be addressed.

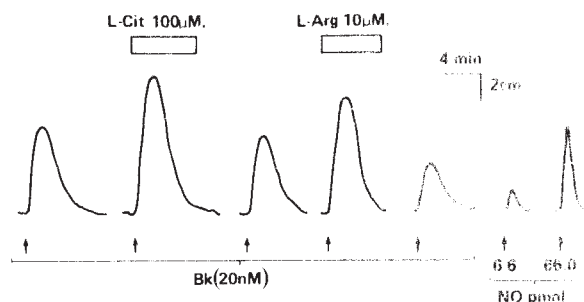


**Fig. 1** The effect of L-arginine and D-arginine on the relaxation of rabbit aortae by NO. A column (1.5 cm in diameter;  $1.2 \times 10^7$  cells) packed with endothelial cells cultured on microcarriers for 10–16 days in Dulbecco's modification of Eagle's medium containing 10% fetal calf serum, penicillin ( $100 \text{ U ml}^{-1}$ ), streptomycin ( $10 \mu\text{g ml}^{-1}$ ) and gentamycin ( $5 \mu\text{g ml}^{-1}$ ) and then for 24 h in culture medium without L-arginine, CM(-Arg) cells, was perfused with Krebs' buffer ( $5 \text{ ml min}^{-1}$ ). The effluent was used to superfuse in a cascade three spiral strips of rabbit aorta denuded of endothelium<sup>2</sup>. The tissues were contracted submaximally by a continuous infusion of 9,11-dideoxy-9 $\alpha$ ,11 $\alpha$ -methano epoxy-prostaglandin  $F_{2\alpha}$  (U46619;  $30 \text{ nM}$ ) and were separated from the cells by delays of 1, 4 and 7 s respectively. The amplification of the recorder was adjusted so that the response of each tissue to glyceryl trinitrate (GTN;  $50 \text{ nM}$ ) administered over the tissues was similar. The bioassay tissues were relaxed by NO ( $44 \text{ nM}$  over the tissues), prepared as described previously<sup>1</sup>; the magnitude of the relaxations declined during passage down the cascade. Nitric oxide, released by a 1 min infusion of bradykinin ( $20 \text{ nM}$ , administered through the column), caused a relaxation of the bioassay tissues which also declined during passage down the cascade. This release was enhanced by infusion of L-arginine ( $10 \mu\text{M}$ ), but not D-arginine ( $10 \mu\text{M}$ ) through the column. Similar results were obtained in three other experiments. RbA, rabbit aorta; Bk, bradykinin.

Infusions of L-arginine ( $1\text{--}10 \mu\text{M}$ ) through a column of cells 5 min before and for 5 min during stimulation of the CM cells with bradykinin ( $20 \text{ nM}$ ) had no effect on the release of NO in 11 out of 16 experiments. In the remainder, the release of NO was increased to a variable extent, ranging from a marginal increase to a maximum of 1.5-fold as measured by chemiluminescence ( $1.24 \pm 0.07$  fold; mean  $\pm$  s.e.m.,  $n = 5$ ). In contrast, the release of NO from CM(-Arg) cells as determined by bioassay (Fig. 1) and chemiluminescence (Fig. 2) was enhanced by L-arginine ( $1\text{--}10 \mu\text{M}$ ). This enhancement was consistent and significant, reaching a maximum of 2.5 fold as measured by chemiluminescence ( $1.9 \pm 0.01$  fold,  $n = 20$ ;  $P < 0.05$ ). The half-maximal effective concentration ( $\text{EC}_{50}$ ) of L-arginine was  $3.5 \pm 0.1 \mu\text{M}$  ( $n = 3$ ). D-arginine ( $10\text{--}100 \mu\text{M}$ ) was without effect on both CM and CM(-Arg) cells ( $n = 3$  for each group of cells).

The release of NO from CM(-Arg) cells induced by bradykinin was also consistently enhanced by L-citrulline ( $10\text{--}100 \mu\text{M}$ ; Fig. 2) to a maximum of 1.76-fold as measured by chemiluminescence ( $1.53 \pm 0.12$ -fold;  $n = 3$ ;  $P < 0.05$ ) with an  $\text{EC}_{50}$  of  $35.1 \pm 2.1 \mu\text{M}$  ( $n = 3$ ). In contrast, L-ornithine, L- $\alpha$ -amino  $\gamma$ -guanidino butyric acid, L-homoarginine, guanidine hydrochloride, canavanine and the dipeptide L-arginyl-aspartate were all without effect ( $100 \mu\text{M}$ ;  $n = 3$  for each compound). This strict structural and isomeric specificity strongly implicates the involvement of an enzyme in the generation of NO from L-arginine.

The response of the bioassay tissues to exogenous NO ( $44 \text{ nM}$ ), and to endogenous NO released by bradykinin was



**Fig. 2** The effect of L-citrulline and L-arginine on the release of NO as determined by chemiluminescence<sup>1</sup>. Effluent of a column of CM(-Arg) cells (see Fig. 1) was passed continuously ( $5 \text{ ml min}^{-1}$ ) into a reaction vessel containing  $75 \text{ ml}$   $1.0\%$  sodium iodide in glacial acetic acid under reflux. NO was removed from the refluxing mixture under reduced pressure in a stream of  $\text{N}_2$ , mixed with ozone and the chemiluminescent product measured with a photomultiplier and quantified by reference to 1 min infusions of authentic NO<sup>1</sup> into the reflux vessel. The release of NO was enhanced by infusions of L-arginine ( $10 \mu\text{M}$ ) and L-citrulline ( $100 \mu\text{M}$ ) through the column. Similar results were obtained in two other experiments. Bk, bradykinin.

not affected by infusions of L-arginine ( $100 \mu\text{M}$ ) or L-citrulline ( $100 \mu\text{M}$ ) administered over the tissues ( $n = 3$  for each compound). Thus these compounds, unlike superoxide dismutase<sup>7,8</sup>, do not enhance the response of the bioassay tissues by preventing the inactivation of NO.

Nitric oxide was also released from CM(-Arg) cells by the calcium ionophore A23187 ( $3 \mu\text{M}$ ; Fig. 3). This release, unlike that induced by bradykinin, was protracted. Infusions of L-arginine ( $10 \mu\text{M}$ ) during this release caused an immediate increase in the generation of NO which was rapidly reversed on termination of the infusion. In contrast, L-citrulline ( $100 \mu\text{M}$ ), under these conditions, produced only a slow and partial reversal of the decline in the release of NO, an effect which was slow to disappear on termination of the infusion. This is consistent with conversion of L-citrulline to L-arginine before its metabolism to NO. Furthermore, the rapid onset and disappearance of the L-arginine-induced enhancement of NO release suggests that the cells use exogenous L-arginine directly for the generation of NO.

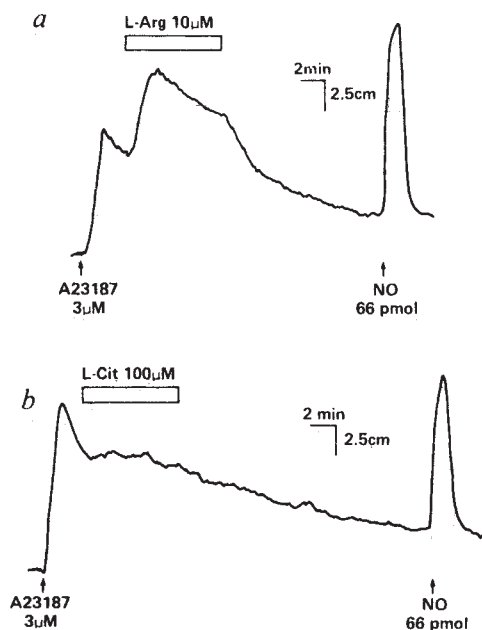
The release of prostacyclin induced by bradykinin ( $100 \text{ nM}$ ) from  $100 \mu\text{l}$  of CM or CM(-Arg) cells ( $2 \times 10^6 \text{ cells ml}^{-1}$ ; 10 min incubation) was similar ( $10.3 \pm 1.1 \text{ ng } 10^{-6} \text{ cells}$  and  $11.7 \pm 1.7 \text{ ng } 10^{-6} \text{ cells}$  respectively;  $P > 0.05$ ) and was not significantly affected by L- or D-arginine ( $100 \mu\text{M}$ ;  $P < 0.05$ ,  $n = 3$ ). These data indicate that L-arginine specifically affects the NO-generating system.

The enhancement of NO release by L-arginine ( $10 \mu\text{M}$ ) could always be demonstrated even when the release of NO had substantially declined after repeated challenges with bradykinin (Fig. 1). This suggests that this decline is due to a failure in the transduction mechanism which follows bradykinin stimulation rather than a failure of the mechanism responsible for the formation of NO.

Studies using  $^{15}\text{N}$ -labelled L-arginine clearly demonstrated that L-arginine is a substrate for the generation of NO (Fig. 4). The release of  $^{15}\text{NO}$  occurred over a similar time course and in amounts consistent with the extent of enhancement of NO release by L-arginine ( $10 \mu\text{M}$ ) as measured by chemiluminescence. The release of  $^{15}\text{NO}$  was similar with both L-(U)- $^{15}\text{N}$ -(95%)-arginine ( $\text{U}^{15}\text{N-Arg}$ ) and L-guanidino- $^{15}\text{N}$ -(99%)-arginine ( $\text{G}^{15}\text{N-Arg}$ ), indicating that only the terminal guanidino nitrogen atom(s) give rise to NO.

Our results suggest that L-arginine is the endogenous substrate for the generation of NO in vascular endothelial cells. Furthermore they point to the existence of two distinct steps in this





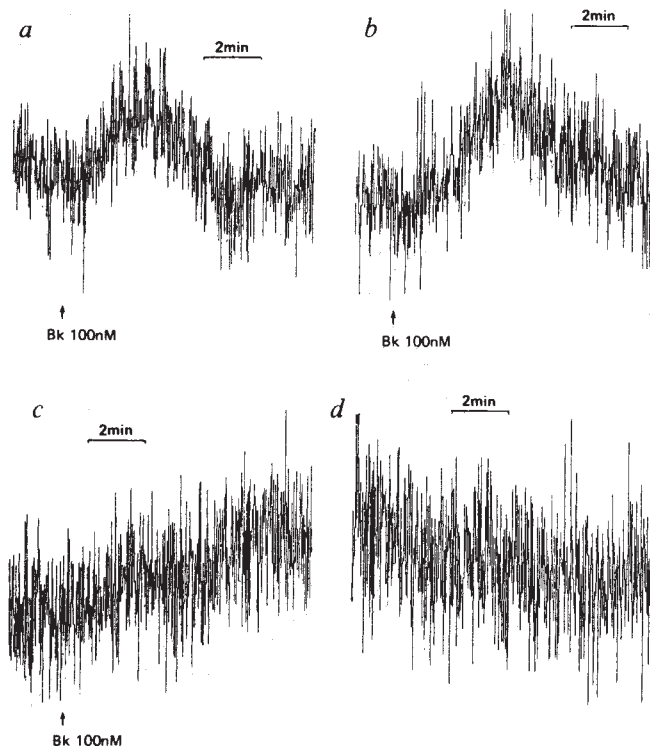
**Fig. 3** The effect of L-arginine and L-citrulline on the release of NO as determined by chemiluminescence. The release of NO from a column of CM(-Arg) cells was determined by chemiluminescence as described in Fig. 2. The calcium ionophore A23187 (3  $\mu$ M), administered as a 1 min infusion, caused a protracted release of NO. When L-arginine (10  $\mu$ M) was infused during the course of the decline in this release there was an immediate increase in NO generation which disappeared rapidly on terminating the infusion (a). When L-citrulline (100  $\mu$ M) was infused under the same conditions there was a reduction in the rate of decline of the release of NO. This effect was slow in onset and was also slow to disappear on termination of the infusion (b). Similar results were obtained in three other experiments.

pathway, firstly, the mobilization of the substrate (L-arginine) and secondly its conversion to NO. Both of these are activated during stimulation of the cells by bradykinin and the calcium ionophore A23187. Changes in the availability of endogenous substrate are sufficient to allow the NO-generating enzyme to use exogenous substrate more effectively. The mechanisms controlling the release of L-arginine and the activation of the NO-forming enzyme require elucidation.

The macrophage is the only other eukaryotic cell reported to form oxides of nitrogen ( $\text{NO}_2^-$ ,  $\text{NO}_3^-$ )<sup>9</sup>. These are derived from the guanidino nitrogen atom(s) of L-arginine and are involved in macrophage cytotoxic activity<sup>10,11</sup>. Unlike endothelial cells, L-homoarginine and L-arginyl-aspartate are substrates and canavanine is an inhibitor of this reaction in macrophages<sup>11,12</sup>. This may reflect differences in experimental conditions or in substrate specificity of the two cell types.

The precise mechanism whereby NO is formed from the terminal guanidino nitrogen atom(s) of L-arginine is not known. It is possible that deimination leads to the formation of  $\text{NH}_3$ , which is then oxidized to NO, as has been suggested for the macrophage<sup>10</sup>. Alternatively, the terminal guanidino nitrogen atom(s) of L-arginine may be N-oxidized in a reaction involving enzyme-mediated attack by oxidizing species such as oxygen radicals or lipid peroxides, generated as a result of the activation of phospholipases.

The importance of L-arginine in the generation of NO by vascular endothelial cells opens up interesting possibilities for the study of the synthesis of this mediator in a variety of pathological states including hypertension, atherosclerosis and vasospasm. Whether NO is released by other cells or has other biological actions is not yet known. The elucidation of its biosynthetic route provides the opportunity to investigate the biological



**Fig. 4** Determination of  $^{15}\text{NO}$  release from endothelial cells by mass spectrometry. A column (5 cm in diameter;  $3.0\text{--}5.0 \times 10^8$  cells) of CM(-Arg) cells was perfused with Krebs' buffer (5 ml  $\text{min}^{-1}$ ) and the effluent passed continuously into the chemiluminescence reflux vessel, as described in Fig. 2, except that a stream of He instead of  $\text{N}_2$  was used to reduce source suppression by the carrier gas. A larger diameter column of cells was used in these studies to increase the amount of NO released and thus to overcome limitations on detection due to the small amount of the He stream that could be sampled by the mass spectrometer source ( $\sim 1\%$ ). The He stream was passed continuously into the source of an MS 50 TC (Kratos) mass spectrometer and  $^{15}\text{NO}$  was determined ( $m/z = 30.9950$ ) by electron impact mass spectrometry with single ion monitoring. The identity of the signal was confirmed by reference to the natural abundance of  $^{15}\text{NO}$  in NO standards which was found to be 0.37%.  $^{15}\text{NO}$  was released from cells stimulated with bradykinin (100 nM) in the presence of an infusion of  $\text{U}^{15}\text{N}$ -Arg (10  $\mu\text{M}$ ; a) or  $\text{G}^{15}\text{N}$ -Arg (10  $\mu\text{M}$ ; b). No change in the baseline signal was observed when the cells were stimulated in the presence of an infusion of  $^{14}\text{N}$ -L-arginine (10  $\mu\text{M}$ ; c) or when cells in the presence of an infusion of  $\text{U}^{15}\text{N}$ -Arg (10  $\mu\text{M}$ ) were not stimulated (d). The cells released  $2.2 \pm 0.4$  nmol  $^{15}\text{NO}$  ( $n = 3$ ) in the presence of  $\text{G}^{15}\text{N}$ -Arg and  $2.1 \pm 0.7$  nmol  $^{15}\text{NO}$  ( $n = 4$ ) in the presence of  $\text{U}^{15}\text{N}$ -Arg, determined by mass spectrometry. Replicate columns of CM(-Arg) cells released an additional  $1.4 \pm 0.2$  nmol ( $n = 4$ ) NO in the presence of an infusion of L-arginine (10  $\mu\text{M}$ ), as determined by chemiluminescence.

consequences of biochemical or pharmacological manipulation of its synthesis.

The authors are indebted to Dr C. I. Pogson for helpful discussions and to Miss L. J. Wallis, D. Cooper, N. A. Foxwell, B. C. Sweatman and M. J. Ashton for technical assistance.

Received 29 March; accepted 12 May 1988.

- Palmer, R. M. J., Ferrige, A. G. & Moncada, S. *Nature* **327**, 524-526 (1987).
- Radomski, M. W., Palmer, R. M. J. & Moncada, S. *Br. J. Pharmac.* **92**, 639-646 (1987).
- Radomski, M. W., Palmer, R. M. J. & Moncada, S. *Biochem. biophys. Res. Commun.* **148**, 1482-1489 (1987).
- Furchgott, R. F. & Zawadzki, J. V. *Nature* **288**, 373-376 (1980).
- Gryglewski, R. J., Moncada, S. & Palmer, R. M. J. *Br. J. Pharmac.* **87**, 685-694 (1986).
- Salmon, J. A. *Prostaglandins* **15**, 383-397 (1978).
- Gryglewski, R. J., Palmer, R. M. J. & Moncada, S. *Nature* **320**, 454-456 (1986).
- Rubanyi, G. M. & Vanhoutte, P. M. *Am. J. Physiol.* **250**, H222-H227 (1986).
- Stuehr, D. J. & Marletta, M. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7738-7742 (1985).
- Hibbs, J. B., Taintor, R. R. & Vavrin, Z. *Science* **235**, 473-476 (1987).
- Hibbs, J. B., Vavrin, Z. & Taintor, R. R. *J. Immun.* **136**, 550-565 (1987).
- Iyengar, R., Stuehr, D. J. & Marletta, M. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6369-6373 (1987).

## The PI-linked receptor FcRIII is released on stimulation of neutrophils

Tom W. J. Huizinga\*, C. Ellen van der Schoot\*, Carolina Jost†, Rob Klaassen\*, Marion Kleijer\*, Albert E. G. Kr. von dem Borne\*, Dirk Roos\*‡ & Pedro A. T. Tetteroo\*

\* Central Laboratory of the Netherlands Red Cross Blood Transfusion Service and Laboratory for Experimental and Clinical Immunology, University of Amsterdam, Amsterdam, The Netherlands  
† Laboratory for Electron Microscopy, University of Leiden, Leiden, The Netherlands

Human phagocytic cells express receptors for the constant (Fc) region of immunoglobulin G. Neutrophils carry Fc receptor II (FcRII; CDw32) and FcRIII (CD16)<sup>1-3</sup> which both bind IgG-containing immune complexes, leading to phagocytosis of the complex and activation of the neutrophil<sup>4,5</sup>. We find that patients with paroxysmal nocturnal haemoglobinuria (PNH) have only about 10% of the normal levels of FcRIII on their neutrophils, whereas the expression of FcRII is unaffected. We show that FcRIII is a phosphatidyl inositol (PI)-anchored protein in neutrophils. Analysis of FcRIII expression in cells of PNH patients, known to be deficient in PI-linked proteins, suggests FcRIII is not PI-linked in monocytes. We find that the synthesis of FcRIII in neutrophils from PNH patients appears normal, indicating that the defect lies in the PI linkage. This lipid linkage of the receptor on neutrophils suggests that its release may be important for its function, and indeed FcRIII release was observed on stimulation of neutrophils by an inflammatory bacterial peptide (f-Met-Leu-Phe), suggesting a role for FcRIII shedding in inflammatory reactions. Activation of the PNH neutrophils with IgG-coated latex beads appeared normal (although binding of dimer IgG complexes was reduced), indicating that FcRII, rather than FcRIII, is involved in neutrophil stimulation.

PNH is an acquired clonal disorder of the bone-marrow stem cell that results in defects of the erythroid, monomyeloid and platelet cell-lineages. In such patients, the expression of PI-linked proteins on the surface of blood cells is decreased.<sup>6</sup> We measured the expression of FcRII and FcRIII on neutrophils purified from three different PNH patients and from control donors. The expression of FcRIII on the PNH neutrophils was less than 10% of that of control neutrophils (Table 1). In contrast, the number of FcRII molecules on the PNH cell surface was normal. As a positive control, we also measured the expression of CD14, a protein on human monocytes believed to be PI-linked<sup>7</sup>. As expected, CD14 was strongly diminished on PNH monocytes. All of the monocytes and neutrophils from these three PNH patients were affected, as mixtures of positive and negative cells were not observed.

The binding of dimer IgG complexes to the neutrophils of PNH patients was also strongly diminished (Fig. 1). Earlier observations demonstrated that this binding is FcRIII-dependent, because it was totally inhibited by anti-FcRIII but not by anti-FcRII monoclonal antibodies<sup>1</sup>. Treatment of neutrophils from normal donors with glycosyl-phosphatidyl-inositol-specific phospholipase C (GPI-PLC), an enzyme which disrupts the PI-linkage, reduced the expression of FcRIII on these cells (Fig. 2), while it did not influence the expression of FcRII (results not shown). From these results and the specific deficit of FcRIII on PNH neutrophils, we conclude that FcRIII is PI-anchored on neutrophils.

The PI-linked form of FcRIII suggests that this receptor can be released. This would be consistent with the observed secretion

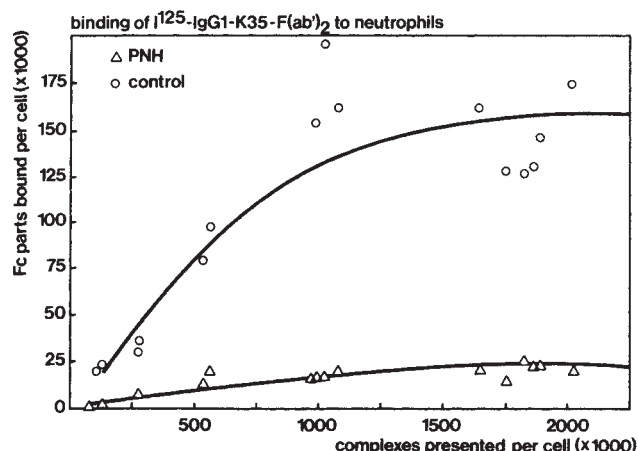


Fig. 1 Binding of different concentrations of <sup>125</sup>I-labelled IgG-K35-F(ab')<sub>2</sub> complexes to neutrophils. One representative experiment shown out of three. All neutrophils in these PNH patients were affected by the disease (see legend to Table 1).

**Methods.** IgG1-K35-F(ab')<sub>2</sub> complexes were made as described<sup>1</sup>. In short, IgG-myeloma proteins were incubated overnight with <sup>125</sup>I-labelled K35 (anti-kappa light chain) F(ab')<sub>2</sub> in a molar ratio of 1:1. Tetrameric complexes are formed consisting of two IgG molecules crosslinked by two F(ab')<sub>2</sub> anti-kappa fragments. These complexes were separated from unbound IgG by Biogel-A5.M gel filtration chromatography. Cells (10<sup>7</sup> ml<sup>-1</sup>) and complexes were incubated on ice for 30 min, layered on a dibutylphthalate/dinonylphthalate oil mixture and spun down to separate unbound complexes from cells with bound complexes. The tubes were frozen and cut to separate the pellet from the supernatant. The pellet and the supernatant were counted separately in a LKB multi-gamma counter. The radioactivity in the pellet was corrected for non-specific binding.

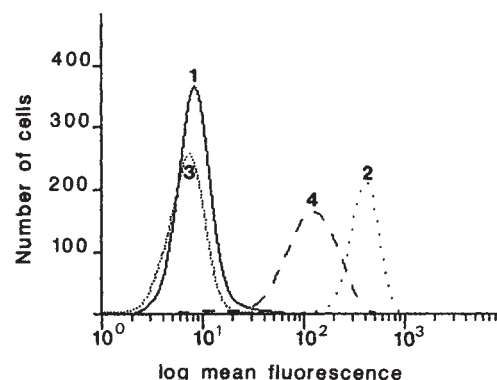


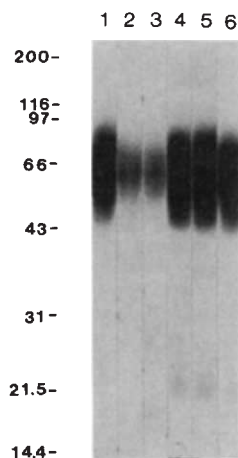
Fig. 2 The expression of CD16 after treatment with the enzyme GPI-PLC, as measured with a flow cytometer. (1) Irrelevant isotype-matched antibody on untreated neutrophils; (2) anti-FcRIII MAb (CLB-FcR-gran1) on untreated neutrophils; (3) irrelevant isotype-matched antibody on GPI-PLC-treated neutrophils; and (4) anti-FcRIII MAb on GPI-PLC-treated neutrophils.

**Methods.** See legend to Table 1. GPI-PLC treatment: neutrophils were isolated and suspended in RPMI-1640 medium (10<sup>7</sup> ml<sup>-1</sup>) and GPI-PLC was added at a final dilution of 1:100 and incubated for 60 min at 37 °C. GPI-PLC was isolated from culture supernatants of *Bacillus thuringiensis* by Dr M. Low.

of an endogenous phospholipase C on activation of cells<sup>8</sup> and reduction in expression of FcRIII after activation<sup>9</sup>. To test this idea, we stimulated neutrophils with the bacterial tripeptide formyl-methionyl-leucyl-phenylalanine (FMLP, a receptor-mediated, non-IgG-dependent activator) and found that FcRIII is indeed released into the supernatant (Fig. 3). In contrast, FcRII was not released (data not shown). Cultured monocytes express FcRIII<sup>10</sup> and are affected by PNH. Cultured PNH monocytes showed expression of FcRIII as control monocytes,

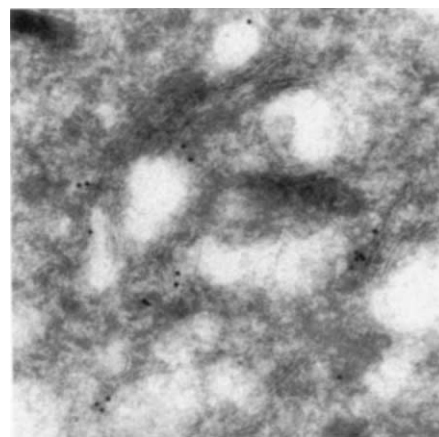
‡ To whom correspondence should be addressed c/o Publication Secretariat, Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, P.P. Box 9406, 1006 AK Amsterdam, The Netherlands.





**Fig. 3** Immunoprecipitation of FcRIII from the culture medium and from cell lysates, and from the culture medium of neutrophils activated with different concentrations of FMLP. One representative experiment out of three is shown. Lane 1, lysate of neutrophils; lane 2, culture medium of non-stimulated neutrophils; lane 3, culture medium of neutrophils stimulated with  $10^{-9}$  M FMLP; lane 4, culture medium of neutrophils stimulated with  $10^{-8}$  M FMLP; lane 5, culture medium of neutrophils stimulated with  $10^{-7}$  M FMLP; and lane 6, culture medium of neutrophils stimulated with  $10^{-6}$  M FMLP.

**Methods.**  $2 \times 10^7$  neutrophils were isolated from a healthy donor and labelled with 37 MBq  $^{125}$ Iodide. After labelling, the cells were suspended in 10 mM phosphate-buffered saline (pH 7.4), supplemented with  $\text{CaCl}_2$  (0.6 mM),  $\text{MgCl}_2$  (1.0 mM), glucose (1 mg  $\text{ml}^{-1}$ ) and human albumin (0.5%, w/v). The cells were treated for 15 min at 37 °C with different concentrations of FMLP and then spun down to separate the culture medium from cells. The cells were lysed in 1% (w/v) NP40 in the presence of the protease inhibitors EDTA, phenylmethylsulphonylfluoride and trypsin inhibitor. FcRIII was immunoprecipitated by anti-FcRIII (CLB-FcR-gran1) Sepharose from the lysates of washed cells or from the 12 000 g supernatants of the cells. The anti-FcRIII Sepharose was subjected to non-reducing 10% SDS-PAGE and autoradiography.



**Fig. 4** Electron micrograph of FcRIII present in the Golgi apparatus of a neutrophilic granulocyte of a PNH patient.

**Methods.** Leukocytes were obtained from heparinized blood and erythrocytes were removed by dextran sedimentation. Cells were fixed in 8% formaldehyde and prepared for cryosectioning<sup>21</sup>. FcRIII was visualized by immunogold labelling of cryosections<sup>22</sup> using CLB-FcR-gran1. Magnification:  $\times 78,000$ .

The intracellular localization of FcRIII in PNH neutrophils was examined by electron microscopy. FcRIII was found to be present in the Golgi apparatus, but few FcRIII were visible on the plasma membrane (Fig. 4). Control neutrophils in contrast show a high expression of FcRIII on the plasma membrane (C. Jost, manuscript in preparation). This suggested that the synthesis of FcRIII in PNH neutrophils is normal, and that it is either the coupling to, or the phospho-inositol anchor itself, that is defective. This defect was also suggested by studies on two other proteins deficient in PNH, acetylcholine esterase and Decay Accelerating Factor<sup>6</sup>.

Neutrophils from PNH patients showed a normal IgG-mediated activation as measured by the oxygen consumption induced by IgG-coated latex. PNH ( $n = 3$ ) as well as control cells ( $n = 7$ ) consumed  $2.5 \pm 0.5$  nmol  $\text{O}_2$  per  $10^6$  neutrophils per min. These results are in accord with unpublished experiments on normal neutrophils which suggested that the major FcR for activation is FcRII, whereas FcRIII mediated the adherence of small IgG complexes.

Two common functions of PI-linked proteins have been suggested. First, PI-linked proteins have high lateral mobility in the membrane because the anchor is entirely contained in the outer leaflet of the phospholipid bilayer, and thus the protein cannot be attached to the cytoskeleton<sup>11</sup>. Such high mobility of the FcRIII would facilitate the adherence of IgG-containing immune complexes to neutrophils by rapid diffusion of this receptor to the site of adhesion. Consistent with this idea, the FcRIII is important for adhesion, comparable to other PI-linked adhesion molecules (leukocyte function-associated antigen 3, neural cell adhesion molecule). Second, PI-linked proteins can be released from the cell surface by hydrolysis of the PI-anchor<sup>12</sup>. Thus, FcRIII could be released upon activation as observed.

The role of FcRIII shedding after activation is not clear. In serum, IgG-binding factors or soluble Fc receptors have been identified<sup>13-15</sup>. These factors are believed to play a role in the regulation of synthesis of IgG by B lymphocytes<sup>16</sup> and may control or modify Fc-mediated functions in immunological/inflammatory reactions<sup>17</sup>. Simmons and Seed<sup>18</sup> found a PI-linkage of FcRIII on lymphocytes. Our results indicate that FcRIII on neutrophils is also PI-linked. Differences between FcRIII on lymphocytes, neutrophils and cultured monocytes were reported previously<sup>3</sup>. We have shown that, on neutrophils, FcRIII is a PI-anchored adhesion molecule that can be released upon activation.

**Table 1** Expression of FcRIII and FcRII on resting human neutrophils

|               | Control cells |     |     | Patient cells |    |     |
|---------------|---------------|-----|-----|---------------|----|-----|
|               | 1             | 2   | 3   | 1             | 2  | 3   |
| FcRIII (CD16) | 921           | 966 | 610 | 63            | 52 | 66  |
| FcRII (CDw32) | 58            | 54  | 31  | 55            | 60 | 145 |
| CD14          | 365           | 272 | 262 | 38            | 0  | 9   |

The expression of CD14 was measured on resting monocytes. The cells were isolated from the blood of 3 PNH patients and 3 control donors. Mixtures of positive and negative cells were not detected with either cell type, so all myeloid cells were affected in the PNH patients. Expression of membrane proteins was measured by incubating  $10^6$  cells with appropriate dilutions of monoclonal antibodies (MAb). FcRIII was detected with MAb CLB-FcR-gran1<sup>3</sup>, FcRII with MAb C1KM5<sup>19</sup> and CD14 with MAb CLB-mon1<sup>20</sup>. The cells were washed, and fluorescein isothiocyanate (FITC)-labelled goat-anti-mouse Ig was added. Fluorescence was measured with a FACSTAR fluorocytometer (Becton & Dickinson). The data represent the mean fluorescence on a linear scale after subtraction of the mean fluorescence of granulocytes incubated with an irrelevant isotype-matched antibody.

whereas the expression of CD14 (known to be PI-linked) on the cultured monocytes of the PNH patients was strongly reduced (mean fluorescence of the control monocytes after culturing for 3 days: CD14, 743, FcRIII, 250; PNH monocytes: CD14: 55, FcRIII: 120). This indicates that monocytes express a transmembrane form of FcRIII in contrast with a PI-linked form in neutrophils.

We thank Prof. Dr R. Goudsmit, Dr J. Ploem and Dr K. Nieuwenhuis for providing us with blood of PNH patients. We also thank Dr M. Low for providing us with GPI-PLC.

Received 28 April; accepted 16 May 1988.

1. Huizinga, T. W. J. *et al.* *J. Immun.* (submitted).
2. Looney, R. J. *et al.* *J. exp. Med.* **163**, 826-836 (1986).
3. Werner, G. *et al.* *Leucocyte Typing II* Vol. 3 (eds Reinherz, P. C., Reinherz, E. L., Haynes, B. F., Nadler, L. M. & Bernstein, I. D.) 109-121 (Springer-Verlag, New York, 1985).
4. Klebanoff, S. J. in *Advances in Host Defence Mechanism* Vol. 1 (eds Gallin, J. L. & Fauci, A. S.) 111-162 (Raven Press, New York, 1982).
5. Henson, P. M., Johnson, H. B. & Spiegelberg, H. L. *J. Immun.* **109**, 1182-1192 (1972).
6. Selvaraj, P., Dustin, M. L., Silber, R., Low, M. G. & Springer, T. A. *J. exp. Med.* **166**, 1011-1025 (1987).
7. Haziot, A., Low, M. G. & Goyert, S. M. *Blood* **70** (suppl.), 93 (1987).
8. Low, M. G., Fluterman, A. H., Ferguson, M. A. J. & Silman, I. *Trends biochem. Sci.* **11**, 212-215 (1986).
9. Tetteroo, P. A. T., Bos, M. J. E., Visser, F. J. & von dem Borne, A. E. G. Kr. *J. Immun.* **136**, 3427-3432 (1986).
10. Clarkson, S. B. & Ory, P. A. *J. exp. Med.* **167**, 408-417 (1988).
11. Noda, M., Yoon, K., Rodan, G. A. & Koppel, D. E. *J. Cell Biol.* **105**, 1671-1677 (1987).
12. Low, M. G. & Saitel, A. R. *Science* **239**, 268-274 (1988).
13. Kay, N. E. & Douglas, S. D. *Int. Archs. Allergy appl. Immun.* **66**, 131-135 (1981).
14. Molenaar, J. L., van Galen, M., Hannema, A. J., Zeijlmeijer, W. P. & Pondman, K. W. *Eur. J. Immun.* **7**, 230-236 (1977).
15. Papamichail, M., Tsokos, G. & Paraskevas, F. *Immunology* **40**, 47-52 (1980).
16. Fridman, W. H., Rabourdin-Combe, C., Neuport-Sautes, C. & Gislser, R. H. *Immun. Rev.* **56**, 51-88 (1981).
17. Limb, G. A., Brown, K. A., Wolstencroft, R. A. & Dumonde, D. C. *Clin. exp. Immun.* **71**, 362-367 (1987).
18. Simmons, D. & Seed, B. *Nature* **333**, 568-570 (1988).
19. Tetteroo, P. A. T., van der Schoot, C. E., Visser, F. J., Bos, M. J. E. & von dem Borne, A. E. G. Kr. in *Leucocyte Typing III. White Cell Differentiation Antigens* (eds McMichael, A. J. *et al.*) 702-706 (Oxford University Press, Oxford, 1987).
20. van der Schoot, C. E., Visser, F. J., Bos, M. J. E., Tetteroo, P. A. T. & von dem Borne, A. E. G. Kr. in *Leucocyte Typing III. White Cell Differentiation Antigens* (eds McMichael, A. J. *et al.*) 603-611 (Oxford University Press, Oxford, 1987).
21. Tokuyasu, K. T. *J. Microscopy* **143**, 139-149 (1986).
22. Slot, J. W. & Geuze, H. J. *Eur. J. Cell Biol.* **38**, 87-93 (1985).

## A new retinoic acid receptor identified from a hepatocellular carcinoma

Doris Benbrook\*, Elisabeth Lernhardt & Magnus Pfahl†

Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

Processes as diverse as growth, vision and reproduction depend on the presence of vitamin A and its metabolites (retinoids), but the molecular mechanisms which govern these diverse actions remain unclear (for reviews see refs 1, 2). A crucial advance recently was the isolation of a specific nuclear receptor for retinoic acid<sup>3,4</sup>, one of the physiologically active vitamin A derivatives. This nuclear receptor is a member of the steroid/thyroid hormone receptor family. Our analysis of an uncharacterized member of this class of intracellular receptors, encoded by a complementary DNA clone from a human placental library, has led us to discover a second retinoic acid receptor. This new receptor is expressed at high levels in a number of epithelial-type tissues. The gene for the receptor was first identified in a hepatocellular carcinoma where it surrounds a site of integration of hepatitis B virus<sup>5</sup>. Activation by this virus may play a role in tumour development in liver cells, where it is normally not expressed.

The DNA-binding domain of steroid hormone receptors and related proteins consists of a highly conserved cysteine-rich region that has the potential to form two zinc-binding fingers<sup>6</sup>. Cellular DNA surrounding a reported site of hepatitis B virus (HBV) integration in a human hepatocellular carcinoma has been reported to encode the first half of such a DNA-binding domain<sup>7</sup>. To isolate and characterize the receptor gene that may have been activated by this HBV integration event, we screened human cDNA libraries with a synthetic oligonucleotide homologous to the cellular DNA sequence downstream of the

HBV integration site. A positive clone from a human placental library<sup>7</sup> (clone HBV-1) was sequenced and revealed complete homology between nucleotides 500 and 649 with the putative exon sequence surrounding the HBV integration site<sup>5</sup>. The suggested open reading frame encodes a 448 amino-acid protein of relative molecular mass ( $M_r$ ) 50,000 (50K), that contains the domains typical of members of the steroid/thyroid hormone receptor family. The cysteine-rich DNA-binding domain (C) is connected through a putative hinge region (domain D) with a ligand-binding domain (E) (see Fig. 1). The amino-terminal region is short and contains only the B domain but no A domain. The A domain is typical of steroid hormone receptors but is not found in thyroid hormone receptors<sup>8</sup>. Comparison of the amino-acid sequence encoded by the HBV-1 clone with other members of this receptor family revealed complete identity with the uncharacterized receptor reported by DeThé *et al.*<sup>9</sup> (except for Leu<sub>407</sub> which was reported as Met) and a very high homology with the recently described retinoic acid receptor (RAR)<sup>3,4</sup> (Fig. 1). The highest homology with the RAR is in the DNA-binding domain where 97% of the amino-acid residues are identical. This suggests that both proteins recognize similar if not identical DNA sequences. The DNA-binding domain also shows considerable homology to the human oestrogen receptor (hER)<sup>10,11</sup> (55%) and to the  $\alpha$  and  $\beta$  human thyroid hormone receptors (hTR)<sup>8,12</sup> (55% and 57%, respectively, Fig. 1a).

The hormone-binding domain is 88% homologous to the RAR implying that this novel receptor is also a receptor for a retinoid. Domain D is also highly conserved and shows 74% homology. Conservation of three colinear domains has so far only been observed between the two human thyroid hormone receptors encoded by the *erbA- $\beta$* <sup>12</sup> and the *erbA-T*<sup>8</sup> genes. In that case, an 86% homology in the hormone-binding domain was sufficient to conserve ligand specificity<sup>8</sup>.

To further analyse the receptor, we synthesized the protein *in vitro*. RNA made *in vitro* containing the complete 5'-untranslated region of the HBV-1 clone was efficiently translated in a rabbit reticulocyte lysate system. The largest protein banded slightly below the largest translation product of hTR $\alpha$  on a denaturing polyacrylamide gel and well below the major hER band (see Fig. 2), consistent with the predicted size of the protein encoded by the HBV-1 clone. The HBV-1 RNA was of a single size, and so the smaller proteins seen in the XR lane may represent translation products initiating at AUG codons within the coding sequence.

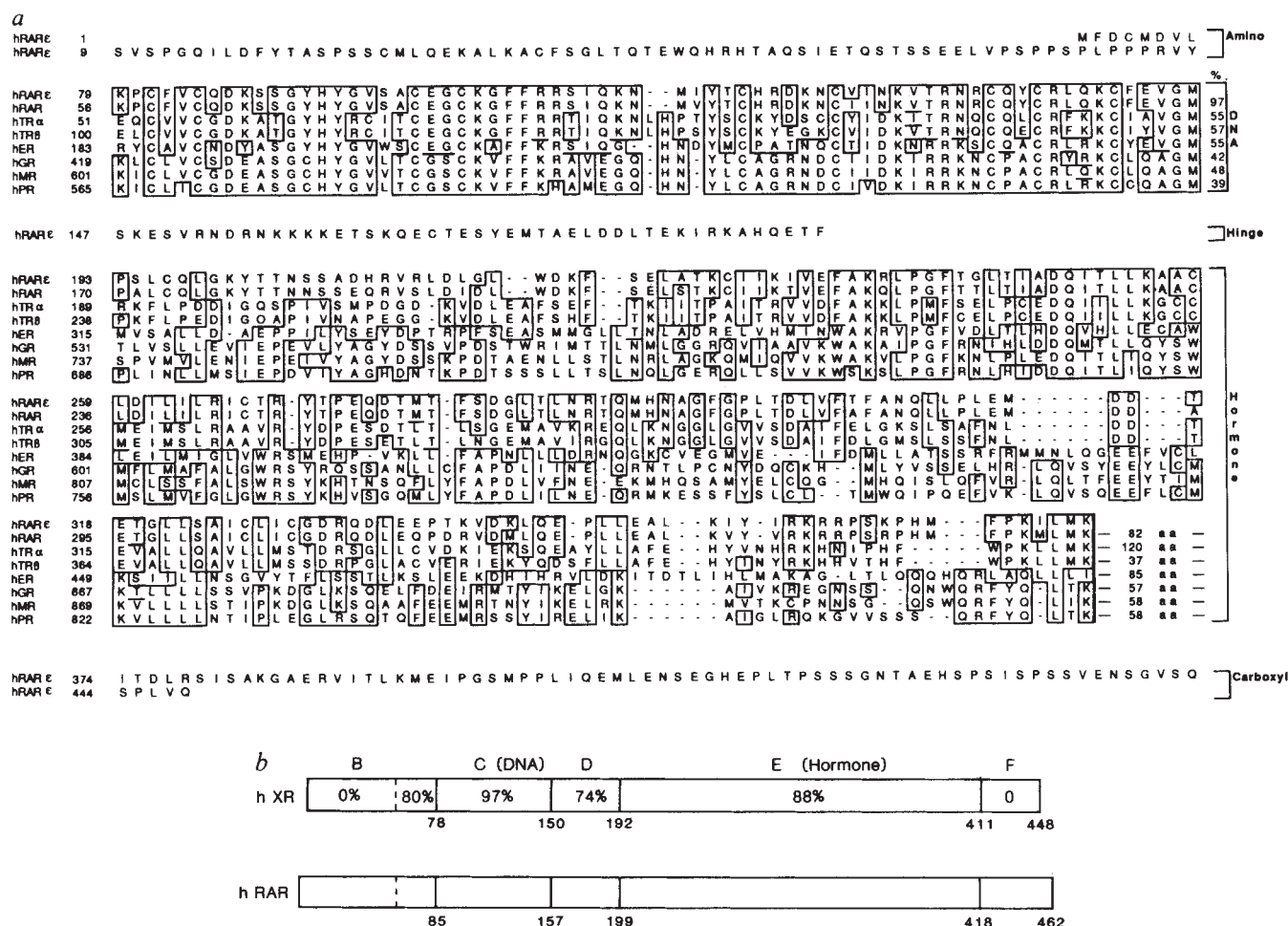
Attempts to measure specific binding of the *in vitro*-synthesized protein to <sup>3</sup>H-labelled retinoids (retinol and retinoic acid) gave inconclusive results. We therefore used the 'finger-swap' approach to determine the ligand-specificity of our novel receptor. In this procedure, pioneered by Green and Chambon<sup>13</sup>, the DNA-binding domain (which presumably forms two zinc-binding fingers) of the unknown receptor is exchanged with the DNA-binding domain of a receptor for which specific hormone-responsive DNA elements (HRE) have been identified. The hybrid receptor is used to transfect susceptible tissue-culture cells together with a reporter gene containing the relevant HRE. The effect of various potential activators of the unknown receptor can then be measured from cells transfected with both genes and grown in the presence and absence of the various ligands<sup>3,4</sup>. We used the hER DNA-binding domain as the donor domain (Fig. 3) and the resulting construct was confirmed by DNA sequencing (data not shown). For further analysis of the construct, RNA of the new hybrid gene was translated *in vitro*. The major band migrated very close to the largest protein made from the wild-type gene, whereas other constructs which did not encode complete receptor proteins gave rise to smaller-sized bands (see Fig. 2b).

Results from transient co-transfection experiments in which we used a CAT (chloramphenicol acetyl transferase) reporter gene linked to a promoter containing the oestrogen-responsive element (ERE) are shown in Fig. 4. Only retinoic acid induces

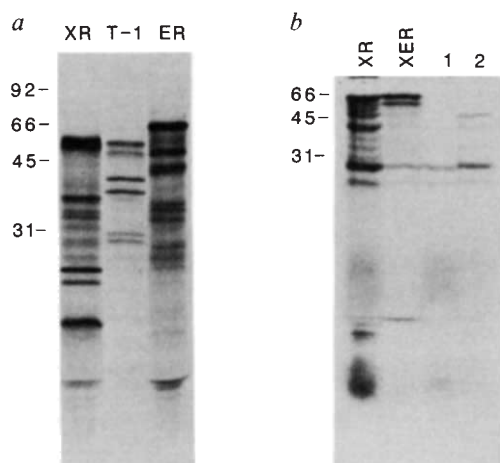
\* Present address: Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Field, London WC2A 3PX, UK.

† To whom correspondence should be addressed.



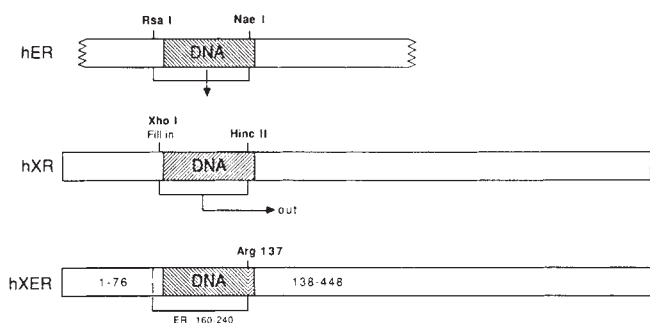


**Fig. 1** Comparison of the amino-acid sequence encoded by the HBV-1 clone with amino-acid sequences of other members of the nuclear receptor family. **a**, The complete amino-acid sequence of the protein (hRAR<sub>E</sub>) encoded by the HBV-1 clone is shown. The sequence was deduced from the nucleotide sequence of a full-length cDNA clone isolated from a human placental library<sup>7</sup>. The amino-acid sequences of the DNA- and hormone-binding domains are compared with the same domains of other human proteins of this family. hRAR, retinoic acid receptor<sup>3,4</sup>; hTR<sub>α</sub> and hTR<sub>β</sub>, thyroid hormone receptors<sup>8,12</sup>; hER, oestrogen receptor<sup>10,11</sup>; hGR, glucocorticoid receptor<sup>14</sup>; hMR, mineralocorticoid receptor<sup>15</sup>; hPR, progesterone receptor<sup>16</sup>. Percentage homologies of RAR<sub>E</sub> with the other receptor proteins in the DNA-binding domain are indicated. Amino-acid residues are boxed when identical in three proteins or more. **b**, Schematic presentation of homologies between the hRAR<sub>E</sub> protein and the hRAR<sup>3,4</sup>. Domains were designated in accordance with Krust *et al.*<sup>17</sup>. The amino-terminal domain A<sup>17</sup> is not present in these proteins. The dotted line in domain B of RAR<sub>E</sub> indicates an exon boundary<sup>5</sup>.



**Fig. 2** Analysis of *in vitro*-synthesized RAR<sub>E</sub> protein. **a**, *In vitro*-transcribed RNAs from clone HBV-1 were used to translate <sup>35</sup>S-labelled RAR protein (lane XR); as controls, human oestrogen receptor (lane ER), and human thyroid hormone receptor α (lane T-1) were synthesized using the same *in vitro* system. **b**, Analysis of the protein from the RAR-ER hybrid gene. An *in vitro* transcription-translation system was used to analyse the protein encoded by the RAR-ER hybrid gene (lane XER) described in Fig. 3. RAR<sub>E</sub> made *in vitro* (lane XR) and translation products from RAR-ER constructs with frameshift mutations (lanes 1 and 2) served as controls.

**Methods.** A 1,704 base pair (bp) *Bam*HI-*Dra*I fragment from clone HBV-1 which contains a 304 bp 5'-untranslated region, the complete open reading frame, and a 55 bp 3'-untranslated region was ligated into the Bluescript vector (Stratagene), cut with *Bam*HI and *Eco*RV (clone B1-RAR). Transcription and translation *in vitro* of the <sup>35</sup>S-labelled RAR<sub>E</sub> protein and the control proteins were performed as described<sup>8</sup>. Proteins were separated on a 15% SDS-acrylamide gel.

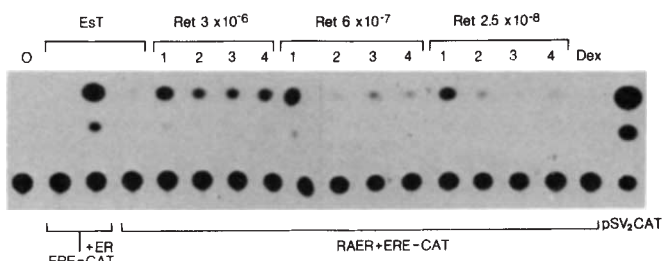


**Fig. 3** 'Finger-swap' from ER to RAR $\alpha$  proteins. Brackets indicate the portions of the RAR $\alpha$  and ER DNA-binding domains used in the construction of an RAR $\alpha$ -ER chimaeric gene. These exchange regions contain the potential zinc-fingers of the DNA-binding domain.

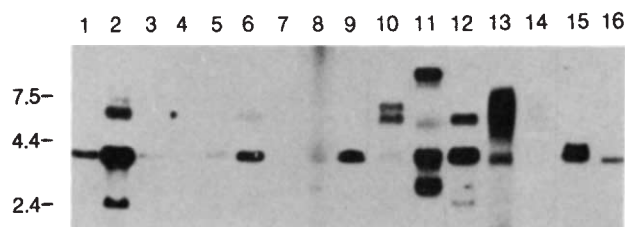
**Methods.** The B1-RAR clone (see legend to Fig. 2) was cut with *Hind*III and *Apa*I, the *Apa*I ends were blunt-ended with mung-bean nuclease and the plasmid was then religated to give a construct (RAR-9) in which the *Xho*I and *Hinc*II sites of the cloning box were deleted. RAR-9 was digested with *Xho*I and *Hinc*II, which cut at unique sites, and the staggered end of the *Xho*I site was filled in. This resulted in a linearized plasmid from which the coding sequences of amino acids 77 to 136 and the first nucleotide of the Arg 137 codon of the RAR $\alpha$  protein were deleted. An *Rsa*I-*Nae*I fragment obtained from the hER cDNA clone, B1-ER<sup>8</sup>, was ligated into the deleted RAR-9 clone. The B1-ER fragment encodes amino-acid residues 168 to 240 and also contains the first nucleotide of the Arg 241 codon of hER. Arg 137 of RAR and Arg 241 of ER are at the same positions in the two DNA-binding domains. The 5' end of the ER fragment encodes 13 amino acids more than were present on the excised RAR fragment. Constructs were analysed by DNA sequencing. The clone with the described features is designated B1-RAER (XER). The major protein obtained from this clone by *in vitro* transcription-translation is indistinguishable in size from the largest RAR $\alpha$  wild-type translation product (Fig. 2b). To express the RAER chimaeric protein in eukaryotic cells, the RAER gene was cut out with *Bam*HI and *Kpn*I (both sides in the cloning box) and ligated into the *Bgl*III and *Kpn*I sites of the PECE eukaryotic expression vector<sup>18</sup>, to give construct PECE-RAER.

CAT gene expression strongly at a physiological concentration ( $2.5 \times 10^{-8}$  M). The hybrid receptor is inactive in the presence of other ligands (oestradiol and dexamethasone). Other retinoids (retinol, retinal and retinyl-acetate) show some inducing effect but only at non-physiological concentrations ( $3 \times 10^{-6}$  M). These results allow us to conclude that the gene activated by HBV integration in a human hepatocellular carcinoma with similarity to steroid-thyroid hormone receptors encodes a retinoic acid receptor. These findings are distinct from the report by DeThé *et al.*<sup>9</sup> who stated that this protein does not specifically bind retinoic acid. The apparent discrepancy is, however, explained by our findings and those of other laboratories<sup>3,4</sup> that specific binding of retinoic acid to its receptors cannot be carried out with protein made *in vitro*. Our results support the usefulness of the finger-swap approach for the identification of novel members of this intracellular receptor family.

To obtain information on the physiological role of this new receptor we investigated the tissue distribution of its messenger RNA (Fig. 5). We chose rat tissue because this species has been extensively used as a model system to establish tissue-responsiveness to retinoic acid<sup>22</sup>. Strong expression was observed in whole brain, pituitary gland, kidney, colon, uterus, ovary, testis, prostate gland, adrenal gland and eye. The size of the transcript ranges from 2.4 to 9.5 kilobase pairs (kb) (our results and ref. 9). The expression pattern is clearly distinct from the previously reported RAR $\alpha$ <sup>4</sup> and shows (with the exception of the brain) a considerable specificity for epithelial-type tissue including skin



**Fig. 4** The RAER chimaeric protein activates the transcription of an ERE-CAT reporter gene in the presence of retinoic acid. HELA cells were co-transfected with the PECE-RAER expression vector (RAER) and the oestrogen-responsive CAT<sup>19</sup> gene (ERE-CAT) and grown in the presence of either  $5 \times 10^{-9}$  M oestradiol (Est) or  $5 \times 10^{-7}$  M dexamethasone (Dex) or the indicated molar concentrations of various retinoids (Ret): 1, retinoic acid; 2, retinol; 3, retinal; 4, retinyl-acetate. As a control, cells were treated with calcium phosphate or were transfected with the constitutive pSV2CAT vector<sup>20</sup>. As further controls, cells were also transfected with ERE-CAT and ERE-CAT plus the oestrogen receptor (ER) expression vector pckR2ER<sup>10</sup> and grown in the presence of Est. **Methods.** HELA cells ( $\sim 1.5 \times 10^6$  per dish) were transfected using calcium phosphate<sup>21</sup> with 5  $\mu$ g of the expression vectors PECE-RAER or pckR2ER and 20  $\mu$ g of either the ERE-CAT vector or the pSV2CAT vector. After 5 h, cells were shocked with 10% glycerol for 2 min. Eighteen hours after the start of transfection, steroid hormones or retinoids were added to the medium and cells were incubated for a further 24 h before harvesting. CAT activity in cell extracts was determined as described<sup>20</sup>.



**Fig. 5** Tissue-specific expression of RAR $\alpha$ . Northern blot analyses were performed with RNA (20  $\mu$ g per lane) extracted from different rat tissues. Lane 1, brain; lane 2, pituitary; lane 3, atrium; lane 4, lung; lane 5, diaphragm; lane 6, kidney; lane 7, liver; lane 8, spleen; lane 9, colon; lane 10, uterus; lane 11, ovary; lane 12, testis; lane 13, prostate; lane 14, seminal vesicle; lane 15, adrenal; lane 16, eye. Markers on the side are in kb.

**Methods.** The guanidine isothiocyanate method<sup>25</sup> was used to isolate RNA from tissues of adult BK1: (SD) rats. Northern blots were prepared from 1% formaldehyde-agarose gels containing 20  $\mu$ g RNA per lane, prehybridized at 42°C overnight in 50% formamide, 5 $\times$ SSPE (1 $\times$ SSPE is 0.18 M NaCl, 10 mM NaPO<sub>4</sub>, pH 7.7, 1 mM EDTA), 5 $\times$ Denhardt's reagent, 0.1% SDS, and 250  $\mu$ g ml<sup>-1</sup> sheared salmon-sperm DNA. Hybridizations were at 42°C in 50% formamide, 5 $\times$ SSPE, 1 $\times$ Denhardt's reagent, 0.1% SDS, 100  $\mu$ g ml<sup>-1</sup> sheared salmon-sperm DNA, and 20 ng ml<sup>-1</sup> of probe 1 for all lanes except 10, 12, 13, and 14 for which 2 $\times 10^6$  c.p.m. ml<sup>-1</sup> of probe 2 was used. Probe 1 was labelled by nick-translating the whole B1-RAR plasmid (10<sup>8</sup> c.p.m.  $\mu$ g<sup>-1</sup>) and probe 2 consisted of an *Eco*RI-*Sph*I fragment of the 5' end of RAR labelled using random oligonucleotides as primers for the Klenow subunit of DNA polymerase (10<sup>9</sup> c.p.m.  $\mu$ g<sup>-1</sup>)<sup>26</sup>. The blots were washed 3 times in 2 $\times$ SSC and 0.1% SDS at room temperature for 20 min and twice in 5 $\times$ SSPE, 1 $\times$ Denhardt's, and 0.1% SDS at 60°C for 1 h. The blots were placed on X-ray film for 16 h.

(data not shown). This prompted us to designate it RAR $\alpha$ . Epithelial-type tissues are known to be clearly affected by vitamin A deficiencies, and a number (but not all) of such symptoms can be reversed by retinoic acid treatments. RAR $\alpha$  mRNA appears to be present at considerably higher levels in epithelial-type tissue than hTR $\alpha$  and hGR mRNAs (data not shown), suggesting that RAR $\alpha$  is a major retinoic acid response mediator in these tissues.



Our discovery of this gene in a hepatocarcinoma is surprising as vitamin A and other retinoids have been considered to be (and are used therapeutically as) anti-tumour drugs. The RAR $\epsilon$  might, however, contribute to tumour development when expressed erroneously in liver tissue<sup>5,9</sup>, where it is normally silent (Fig. 5). Retinoids are known to maintain the proliferative state of epithelial cells<sup>24</sup>. Thus, this type of cell proliferation, when induced in other types of tissue by erroneous RAR $\epsilon$  expression and the presence of retinoic acid, may lead to tumour development. Studies to investigate the frequency of RAR $\epsilon$  expression in human tumours are now under way.

We thank our colleagues at the La Jolla Cancer Research Foundation for discussions and encouragement, Elizabeth Mather for her critical reading of the manuscript, Gerhart Rytzel for the ERE-CAT vector, Robert Oshima for the PECE vector, Michaela Pfahl for excellent technical help, and Diana Lowe for preparing the manuscript. This work was supported by grants from the NSF and the NIH.

The sequence data will appear in the EMBL/GEN Bank/DBJ Nucleotide Sequence Data bases under accession number X07282.

Received 10 March; accepted 6 May 1988.

- Sporn, M. B. & Roberts, A. B. *Cancer Res.* **43**, 3034-3040 (1983).
- Lotan, R. *Biochim. biophys. Acta* **605**, 31-91 (1980).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624-629 (1987).
- Dejean, A., Bougueleret, L., Greschik, K.-H. & Thiollais, P. *Nature* **322**, 70-72 (1986).
- Green, S. & Chambon, P. *Nature* **324**, 615-619 (1986).
- Millán, J. L. *J. biol. Chem.* **261**, 3112-3115 (1986).
- Benbrook, D. & Pfahl, M. *Science* **238**, 788-791 (1987).
- DeThé, H., Marchio, A., Thiollais, P. & Dejean, A. *Nature* **330**, 667-670 (1987).
- Green, S. *et al. Nature* **320**, 134-139 (1986).
- Greene, G. L. *et al. Science* **231**, 1150-1154 (1986).
- Weinberger, C. *et al. Nature* **324**, 641-649 (1986).
- Green, S. & Chambon, P. *Nature* **325**, 75-78 (1987).
- Weinberger, C., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. *Nature* **318**, 610-612 (1985).
- Arriza, J. L. *et al. Science* **237**, 268-275 (1987).
- Misrahi, M. *et al. Biochem. biophys. Res. Commun.* **143**, 740-748 (1987).
- Krust, A. *et al. EMBO J.* **5**, 891-897 (1986).
- Ellis, L. *et al. Cell* **45**, 721-732 (1986).
- Klein-Hitpass, L., Schorpp, M., Wagner, U. & Rytzel, G. U. *Cell* **46**, 1053-1061 (1986).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
- Wigler, M. *et al. Cell* **11**, 223-232 (1977).
- Omori, M. & Chytil, F. J. *J. biol. Chem.* **254**, 14370-14374 (1982).
- Thompson, J. N. in *The Fat Soluble Vitamins* (eds DeLuca, H. F. & Suttie, J. E.) 267-281 (University of Wisconsin Press, Madison, 1969).
- Shapiro, S. S. in *Retinoid and Cell Differentiation* (ed. Sherman, M. I.) 29-60 (CRC Press, Boca Raton, 1986).
- Maniatis, T., Fritsch, E. F. & Sambrook (eds) in *Molecular Cloning, A Laboratory Manual* 196 (Cold Spring Harbor Laboratory, New York, 1982).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).

## Leukocytes express a novel gene encoding a putative transmembrane protein-kinase devoid of an extracellular domain

Yinon Ben-Neriah & Asne R. Bauskin

The Lautenberg Center for General & Tumor Immunology,  
The Hebrew University-Hadassah Medical School,  
Jerusalem 91010, Israel

**Tyrosine-specific phosphorylation of proteins is a key to the control of diverse pathways leading to cell growth and differentiation<sup>1</sup>. The protein-tyrosine kinases described to date are either transmembrane proteins having an extracellular ligand binding domain<sup>2-7</sup> or cytoplasmic proteins related to the *v-src* oncogene<sup>8-10</sup>. Most of these proteins are expressed in a wide variety of cells and tissues; few are tissue-specific<sup>1</sup>. Previous studies have suggested that lymphokines could mediate haematopoietic cell survival through their action on glucose transport<sup>11,12</sup>, regulated in some cells through the protein-tyrosine kinase activity of the insulin receptor<sup>13</sup>. We**

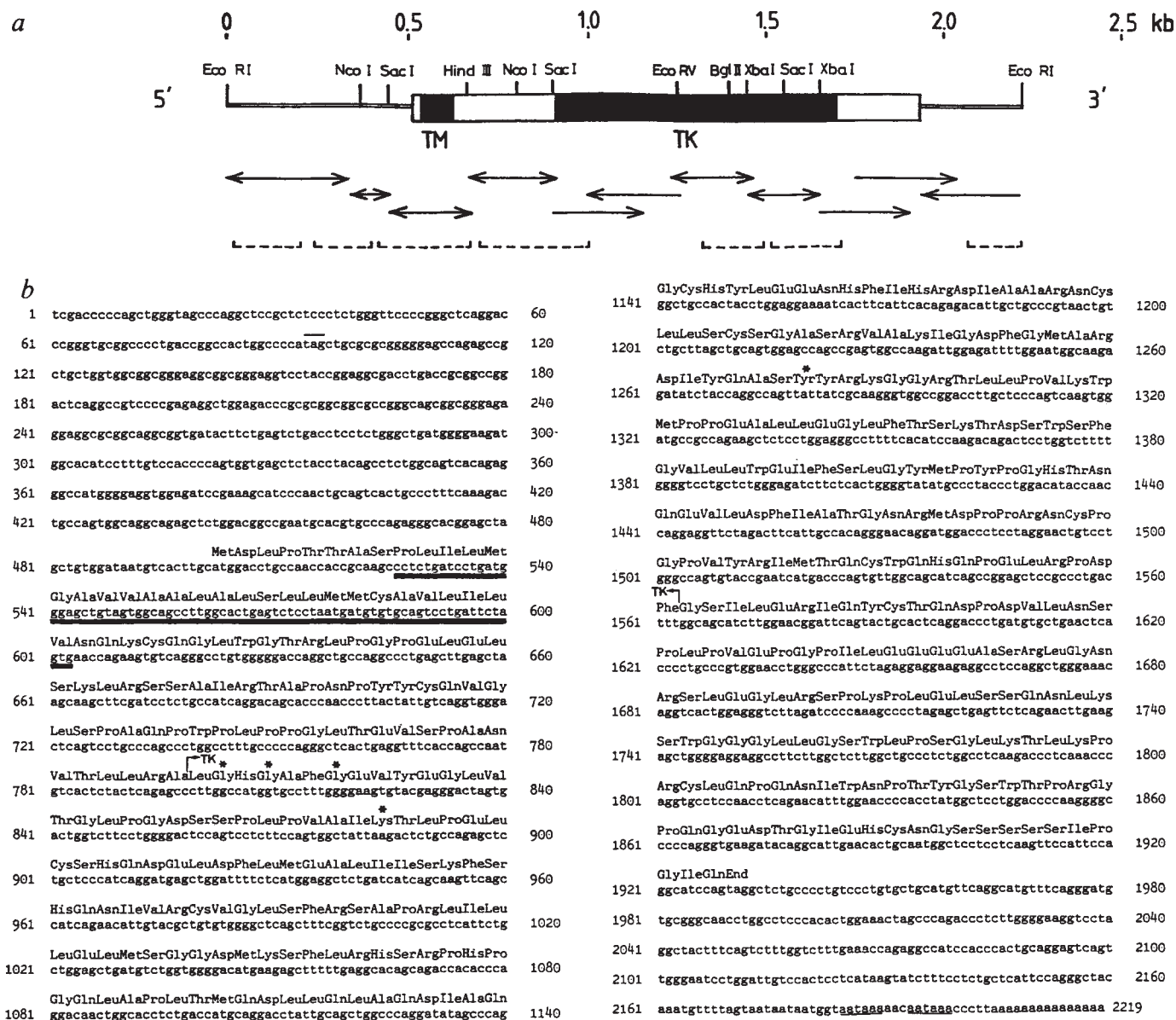
**have investigated the possibility that insulin receptor-like genes are expressed specifically in haematopoietic cells. Using the insulin receptor-related avian sarcoma oncogene *v-ros* as a probe<sup>14</sup>, we have isolated and characterized the complementary DNA of a novel gene, *ltk* (leukocyte tyrosine kinase). The *ltk* gene is expressed mainly in leukocytes, is related to several tyrosine kinase receptor genes of the insulin receptor family and has unique structural properties: it apparently encodes a transmembrane protein devoid of an extracellular domain. Two candidate *ltk* proteins have been identified with antibodies in the mouse thymus, and have properties indicating that they are integral membrane proteins. These features suggest that *ltk* could be a signal transduction subunit for one or several of the haematopoietic receptors.**

The mouse pre-B lymphocyte 70Z/3 cDNA library<sup>10</sup> was screened with a probe derived from the 3' region of avian *v-ros* (ref. 14).  $6 \times 10^5$  recombinant plaques yielded three positive clones sharing similar restriction maps, one of which (pRK14) was further characterized by nucleotide sequencing (Fig. 1). The sequence contains one long open-reading frame of 476 amino acids from the methionine at nucleotide 502 to the termination codon at nucleotide 1,930. The other two reading frames are frequently interrupted by stop codons already present in the first 65 nucleotides. Preceding the first methionine in the same reading frame is an upstream termination codon (nucleotide 94). The coding region is followed by a pair of polyadenylation signals at positions 2,184 and 2,193, and a stretch of poly(A), beginning at position 2,204.

The predicted polypeptide has a relative molecular mass of 52,212 (52.2K) and, when plotted according to the hydrophilicity table of Hopp and Woods<sup>15</sup>, reveals a predominant stretch of 26 hydrophobic amino acids at the N-terminus of the polypeptide (positions 9-34), followed by three polar amino acids. This region, although it has an unusual position, is compatible with the properties of a transmembrane domain and the positively charged peptide (Asn-Gln-Lys) may serve as a membrane transfer stop signal. Following the hydrophobic stretch is a region homologous to the cytoplasmic region of the protein-tyrosine kinases. It has the characteristics of a protein-tyrosine kinase domain, including the consensus sequence Gly-X-Gly-X-Phe-Gly-X-Val (residues 101-108), that together with Lys 128 are thought to create a nucleotide binding site<sup>1</sup>, and the putative autophosphorylation site at Tyr 260.

To determine whether *ltk* is homologous with known proteins, we have searched the GenBank protein data bank and compared *ltk* directly with recently published protein kinase-encoding genes using the computer program GAP (University of Wisconsin Genetics Computer Group). Six proteins, all of them receptors, show more than 45% similarity with *ltk* throughout the entire kinase domain (Fig. 2). These include the human *c-ros* protein (53%) (ref. 16), the *Drosophila* homeotic protein *sev* (49%) (ref. 17), the human receptors for insulin and insulin-like growth factor-I (IGF-1) (48%) (refs 3, 6) and the receptor portion of the human *trk* protein (45%) (ref. 18). Comparison of *ltk* with other kinase-receptor proteins reveals a lower degree of similarity at 30-38%. *ltk* shares a unique peptide (Tyr-Tyr-Arg-Lys-X-Gly-X-X-Leu-Leu-Pro-Val) with the former group of receptor proteins at the autophosphorylation site (Fig. 2). Interestingly, the insulin and IGF-1 receptors are among the few kinases where autophosphorylation activates the kinase activity of the protein<sup>19-21</sup>. It therefore seems that *ltk*, together with the former group of kinase-receptor genes, forms a sub-family within the protein-tyrosine kinase family. It remains to be seen whether the members of this sub-family are also functionally related.

*ltk* is transcribed from a single-copy gene in the mouse genome, and Southern-blot hybridization, using three different restriction enzymes, distinguishes it from *c-ros* (ref. 16) (not shown). The expression of *ltk* was analysed in different murine tissues. Northern blot analysis reveals two transcripts in Rad-LV transformed T-cell lines, Tc38 and Tc23 (ref. 22) (Fig. 3a) and



**Fig. 1** Molecular characterization of a cDNA clone of *ltk*. **a**, Schematic diagram of clone pRK14, showing the *EcoRI* sites that delimit the boundaries of the pRK14 insert. The solid line indicates untranslated sequences; the large box represents coding sequences, in which the putative transmembrane (TM) and tyrosine-specific kinase (TK) domains are indicated. Arrows indicate regions and strands subjected to sequence analysis by the dideoxy chain termination method<sup>29</sup>. The broken line indicates regions also sequenced by the chemical sequencing method<sup>30</sup>. **b**, Nucleotide sequence and the deduced amino acid sequence of pRK14. Nucleotides are numbered on both sides. The putative transmembrane region is underlined with a solid bar. The polyadenylation signals are underlined. The in-frame upstream termination codon is overlined. Asterisks: Gly 101, Gly 103, Gly 106 and Lys 128 indicate the potential ATP binding site, and Tyr 260, the putative autophosphorylation site.

**Methods.** Preparation of mouse 70Z/3 poly(A)<sup>+</sup>RNA and cDNA cloning in  $\lambda$ gt11 was previously described<sup>10</sup>. Screening of recombinants ( $6 \times 10^5$ ) was performed by hybridization of replica nitrocellulose filters using a probe derived from the 3' region of avian *v-ros* (0.8 kb *EcoRI*-*PvuII* fragment)<sup>14</sup>. Filters were hybridized at 38 °C in 50% formamide,  $6 \times$  SSC (0.9 M NaCl, 0.9 M NaCit.),  $5 \times$  Denhardt's solution,  $200 \mu\text{g ml}^{-1}$  denatured herring sperm DNA,  $2 \times 10^7$  c.p.m. of radiolabelled probe and washed in  $2 \times$  SSPE (0.3 M NaCl, 20 mM PO<sub>4</sub>, 2 mM EDTA), 0.1% SDS at 52 °C. Positive clones were subcloned into pSP65 and analysed with restriction endonucleases. Restriction fragments covering the entire cDNA insert of pRK14 were further subcloned into M13 mp18 and mp19 phage vectors (as indicated above) for sequencing by the dideoxy chain termination method<sup>29</sup>. The 5' 760 base-pair region, rich in GC content, was sequenced on both strands using at least one additional method: either by the chemical sequencing method<sup>30</sup> or by a modified chain termination method substituting deoxy-7-deazaguanine triphosphate<sup>31</sup> or dITP<sup>32</sup> for dGTP.

in the thymus (not shown) of ~2.6 and 3.0 kilobases (kb). The expression of *ltk* in the various mouse tissues and cell lines was quantified by RNA slot-blot hybridization analysis using a 5' *ltk* probe (excluding the kinase coding region) (Fig. 3b). *ltk* was abundant in the thymus and observed in the spleen and kidney. Among the different cell lines, *ltk* was expressed in: an interleukin-2 (IL-2)-dependent T-cell line, CTLD4; a macrophage

colony-stimulating factor (M-CSF) dependent macrophage cell line, Mac26; the IL-3-dependent cell line, Ba/F3, thought to be an early pre-B lymphocyte line (R. Palacios, personal communication); and mouse lymphokine-activated killer (LAK) cells. Rehybridization of the blot with an *ltk* kinase-region probe shows a similar tissue distribution, except for an additional faint signal in the brain (not shown). It thus appears that *ltk* is



|               |      |                                                                  |
|---------------|------|------------------------------------------------------------------|
| <i>ltk</i>    | 3    | LPTTASplilmgavvaal.aisilmmcavililvNQKCOGLNGTRL..PGPELELSKLRS     |
| <i>c-ros</i>  | 66   | FWIPetstfilitiivgifivvtipitfwwHRLKNQSAKEGVTVL..INEDKELAEELGL     |
| <i>sev</i>    | 1562 | RGSivlailiapaktvssc.vialvivrKVQRRLRAKKLQOQR..PSIMSNLSLTQTQ       |
| <i>IR</i>     | 892  | AGNSGWEPTFYFYVDYLDVPSNIAKililgplifvif. svvigsyilf1RKRQDPGLG      |
| <i>IGF-IR</i> | 887  | TDPVFFVYQAKTGyenfi.Hilialpvaavililvgglvimilvyfh....RKRNNERLNG    |
| <i>trk</i>    | 237  | Etpfgvavavglvafacil.flatililvINKCGRNKGINRPVAVL...APEDGLAMSLHF    |
| ★             |      |                                                                  |
| <i>ltk</i>    | 60   | AIRTA.P.NPYVCQVGLSPAQWPLPPCLTEVSPANV.....TLLRALGHGAFGEVYEG..LVT  |
| <i>c-ros</i>  | 124  | AAGVGLANACYAINTL.PTQ..EEIENLPAPFREKL.....TLRLLLGSGAFGEVYEGTAVD   |
| <i>sev</i>    | 1618 | QQLMAV.RNRAFSTTLSDADIALLPQINWSQL.....KLLRLFGSGAFGEVYEG..QL       |
| <i>IR</i>     | 991  | PLYASS.NPEVLSASDDVFPSCVVPDEWEVSREKI.....TLLRELQGSFGMVYEGNA..R    |
| <i>IGF-IR</i> | 941  | VLYASV.NPEVFSAAD.....VYVPDEWEVAREKI.....TMSRELQGSFGMVYEGVA..K    |
| <i>trk</i>    | 293  | MTLGG.SLSPTTEKGSGGLGHIIIDPQYFSDACVHHIKRDIIVLKWELGEGAFGVFLA.ECH   |
| ★             |      |                                                                  |
| <i>ltk</i>    | 115  | GLPGDSSPLVAIKTLPELCSHQEDLFLMEALIIKSFHONIVRCVGLSFRSAPRLIILLEMS    |
| <i>c-ros</i>  | 178  | ILGVGSGEIKVAVKTLKKGSTDOEKIEFLKEANLMSKPNILKOLGVCLNEPOYIILEME      |
| <i>sev</i>    | 1668 | KTEDESEFORVAIKELRKGAEEFAEL..LOEAOLMSNPNENIVRLVGICFDTESISLIMBRE   |
| <i>IR</i>     | 1006 | DIKGEAETRVAVKTVNESASLRERIEFLNEASVMKQFTCHHVRLGVVSKGQFTLVVMEMLA    |
| <i>IGF-IR</i> | 990  | GVVDEPETRVAIKTVNEAASMRERIEFLNEASVMKEFNCHHVRLGVVSKGQFTLVIMELMT    |
| <i>trk</i>    | 365  | NLLPEQDKMLVAVKALKE.ASESARODFOREAELLTMLQONIVRFGVCTEGRLMVFEYMR     |
| ★             |      |                                                                  |
| <i>ltk</i>    | 179  | GGDKMSFLRHSP...HPGOLA.....PLTHQDLLQLAQDIAQGGHYLENNFIHRDIAA       |
| <i>c-ros</i>  | 242  | GGDLLTYLRKARM.....ATFYGP.....LLYLVLDLVDLCVDISKGVYLERMFPIHRDLAA   |
| <i>sev</i>    | 1730 | AGDLLSYLRAARATSTQEPPTA.....GLSLSELANCIQVANGSYLEDMMFVHRDLAC       |
| <i>IR</i>     | 1070 | HGDLKSYLRSRPEAENNGPRP.....PPTLQENIQMAEADGMAYLNAKKPVHRDLAA        |
| <i>IGF-IR</i> | 1054 | RGDLKSYLRSRPEAENNGPRP.....PPTLQENIQMAEADGMAYLNAKKPVHRDLAA        |
| <i>trk</i>    | 418  | HGDLNRRFLRSHGP...DAKLLAGGEDVAPGFLQGLQGLAVASQVAGMAYVLAGLHFVHRDLAT |
| ★             |      |                                                                  |
| <i>ltk</i>    | 231  | RNCLLSGSGAS...RYAKIGDFGMARDIYQASYRRKGGRTLLPVKMPPEALLEGFTSKTDS    |
| <i>c-ros</i>  | 294  | RNCLVSEKDYT..SPRIYKIGDFGLARDIYKNDYRKRGEGLLPVRNMAPESLMDGIFTTQSDV  |
| <i>sev</i>    | 1786 | RNCLVTESTGSTRRTYKIGDFGLARDIYKSDYRKRGEGLLPVRNMAPESLMDGIFTTQSDV    |
| <i>IR</i>     | 1125 | RNCMVADFTV.....KIGDFGMTRDIYETDYRRKGGKGLLPVRNMAPESLMDGIFTTSSDM    |
| <i>IGF-IR</i> | 1109 | RNCMVADFTV.....KIGDFGMTRDIYETDYRRKGGKGLLPVRNMAPESLMDGIFTTSSDV    |
| <i>trk</i>    | 478  | RNCLV...GGG...LVYKIGDFGMSRDIYSTDYRKGGRTHLIRNMPRESILYRKETTESDV    |
| ★             |      |                                                                  |
| <i>ltk</i>    | 291  | WSFGVLLWEIFSLGYMPPGHIMQEVLDFIATONRMDPPRNCRGPVYRINTQCNQHOPELRPDF  |
| <i>c-ros</i>  | 356  | WSFGVLLWEILTLGHOPYRANSLDVLNVYQTGGRLPPRNCRDLLNIMTQCAQEPDORPTF     |
| <i>sev</i>    | 1860 | WAFGVLCNEILTLOOPYAARNNFVLAHVKEGDLQPPHCKEKLVSLLLCLWRDTPWERPSF     |
| <i>IR</i>     | 1182 | WSFGVLLWEITSLAEOPYOGLSNEQVILFVMDGGYLDQPCNCPERTDILRMCMQFNPKMRPTF  |
| <i>IGF-IR</i> | 1166 | WSFGVLLWEITSLAEOPYOGLSNEQVILFVMEGLLDKPCNCPDMLFELRMCMQFNPKMRPTF   |
| <i>trk</i>    | 535  | WSFGVLLWEIFTYQKQPMYQLSNTEAIDCITQRELERPRACPFVYVAINRGCVOREPSNATAS  |

**Fig. 2** The insulin receptor family: comparison of the amino acid sequence of the putative *ltk* protein with the corresponding domains of the human *c-ros* protein<sup>16</sup>, the *Drosophila sev* protein<sup>17</sup>, the human insulin receptor (HIR) (refs 2 and 3), the human insulin-like growth factor receptor (IGF-1-R) (ref. 6) and the human *trk* protein<sup>18</sup>. Dots are gaps introduced to optimize alignment. Amino acids constituting the putative transmembrane sequences are indicated by lowercase letters. Matching residues are boxed and shaded. The putative sites for nucleotide binding and autophosphorylation are indicated by asterisks.

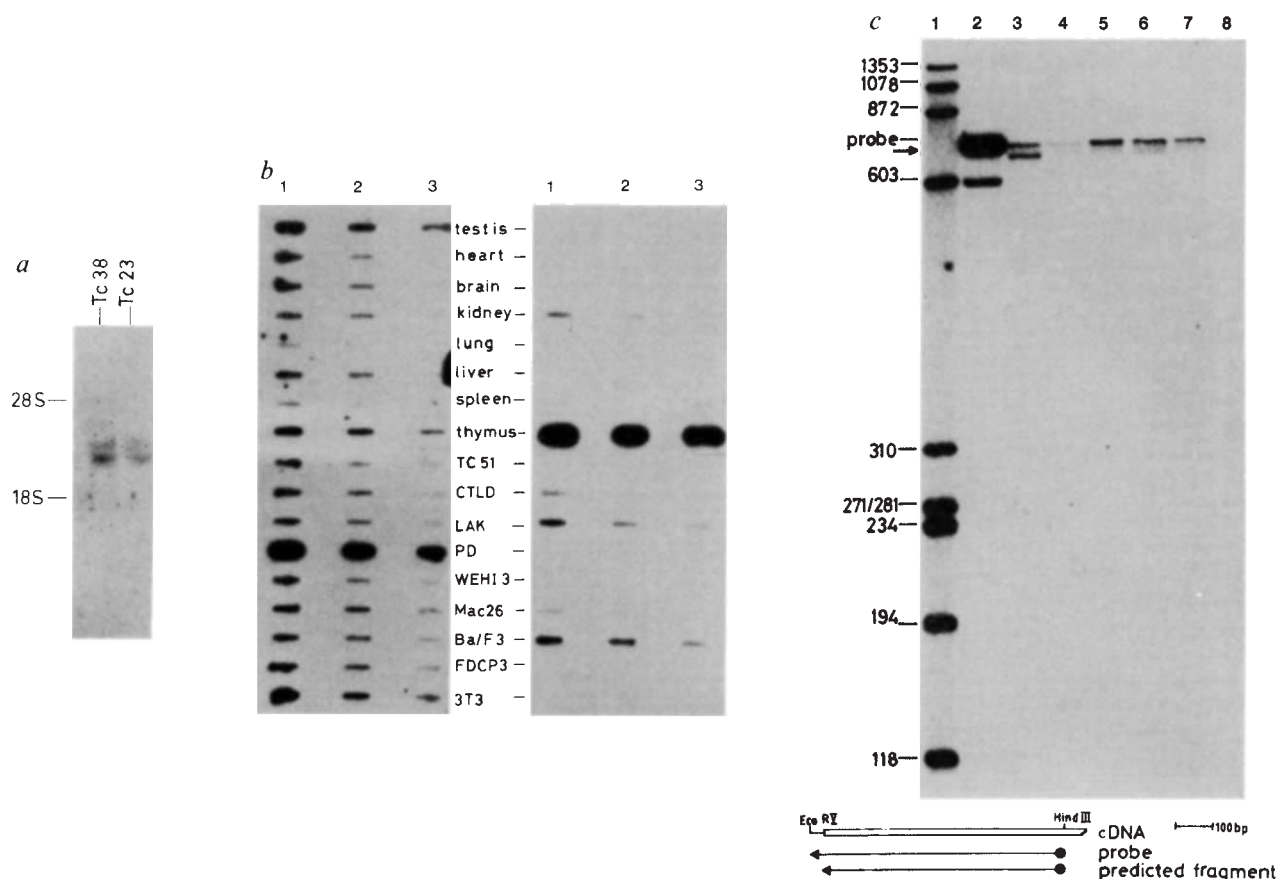
expressed mainly in cells of a haematopoietic origin, particularly in cells of the T lineage.

Considering the homology of *ltk* with transmembrane protein-kinases, it was expected to include an extracellular recognition unit. But the putative external portion of the protein is only eight amino acids long and unlikely to be sufficient for ligand binding. It is possible that one of the two transcripts detected by the *ltk* probe (Fig. 3a) encodes another protein with a larger extracytoplasmic domain. N-terminal variation has been observed in several other protein-tyrosine kinases as a result of alternative messenger RNA splicing<sup>10,23</sup>. The *c-abl* proto-oncogene generates two mRNA transcripts of 5.5 and 6.5 kb by alternative splicing, each encoding a protein with a different N-terminus<sup>10</sup>. If the origin of the two *ltk* transcripts is similar, we may have characterized one alternative splice form, whereas the other may encode a different N-terminus. Nevertheless, RNA hybridization studies demonstrate the expression of transcripts homologous to the 5' region of *ltk* in several tissues and cell lines (Fig. 3b). To verify that these transcripts indeed contain the 5' coding sequence predicted from the cDNA clone pRK14, the following expression studies were performed: radiolabelled probe corresponding to the 5' region of *ltk* cDNA was hybridized to total cellular RNA from various tissues and subjected to S1 nuclease digestion. The only band observed (Fig. 3c) corresponds in size to the fully protected probe, indicating the presence in thymus of an *ltk* transcript fully complementary to the 5' region of pRK14 cDNA.

To demonstrate that *ltk* encodes a transmembrane protein,

*ltk*-specific antiserum was prepared against an *ltk*/β-galactosidase fusion protein and used to demonstrate the presence of *ltk* protein in a thymus membrane protein fraction (Fig. 4). The specificity of the serum was shown by immunoprecipitation of *ltk* protein synthesized *in vitro*. The entire *ltk* cDNA was transcribed *in vitro* and the transcript used for protein synthesis in rabbit reticulocyte lysate (Fig. 4a, lane 2). The larger protein band has a relative molecular mass of 52K, in agreement with the predicted size of the protein encoded by the long open-reading frame of *ltk*. The antiserum directed to the C-terminal portion of the putative *ltk* protein specifically immunoprecipitated the 52K protein along with the smaller protein fragments (Fig. 4a, lane 4), showing that they probably correspond to downstream methionine initiation.

A Triton X-114 detergent-fractionation assay<sup>24</sup> was used to show that the *ltk* protein in thymus is an integral membrane protein. Transmembrane proteins have been shown to partition into the Triton X-114 detergent phase during fractionation, whereas the more hydrophilic cytosolic and secretory proteins remain in the aqueous phase<sup>24,25</sup>. Western blot analysis of Triton X-114 detergent-fractionated thymus proteins showed that the anti-*ltk* serum reacts specifically with two protein bands of 56K and 64K in the Triton detergent phase (Fig. 4b, lane 4). Furthermore, the addition of the thymus detergent-phase proteins to the antiserum inhibits its reaction with the 52K *in vitro* translated *ltk* protein (Fig. 4a, lane 5), whereas no inhibition was observed with the thymus aqueous-phase proteins (Fig. 4a, lane 6). The 56K band is consistent with the expected size of the encoded



**Fig. 3** Expression of *ltk* in different murine tissues and cell lines. *a*, Northern blot analysis of poly(A)<sup>+</sup> RNA from Rad-LV transformed T-cell lines<sup>22</sup>. Poly(A)<sup>+</sup> RNA (5 µg per lane) was hybridized with radiolabelled 1.6 kb *Hind*III-*Eco*RI fragment (see Fig. 1*a*). RNA was extracted by LiCl<sub>2</sub>-urea lysis and phenol/chloroform extraction, followed by oligo(dT) selection and separated on a formaldehyde-containing 1% agarose gel. The RNA was transferred to Zetabind nylon membrane, hybridized at 42 °C in 50% formamide, 6×SSC, 5×Denhardt's solution, 500 µg ml<sup>-1</sup> denatured herring sperm DNA, 2×10<sup>7</sup> c.p.m. of radiolabelled probe and washed at 68 °C in 2×SSPE, 0.1% SDS. Exposure was for one week at -70 °C using an intensifying screen. *b*, Slot-blot analysis of total cellular RNA isolated from different murine tissues and cell lines. RNA (30 µg, lane 1; 10 µg, lane 2; and 3 µg, lane 3) was spotted onto nitrocellulose using the Minifold II slot-blotter (Schleicher and Shuell) and hybridized with a radiolabelled *c-abl* kinase region probe as a control, left panel; and a radiolabelled 670 bp *Eco*RI-*Hind*III fragment (see Fig. 1*a*), right panel. Hybridization and washing conditions are described above. Exposure was for 2 days at -70 °C using an intensifying screen. Cell lines: TC51, Rad-LV transformed T-cell line<sup>22</sup>; CTLD, an IL-2-dependent T-cell line; LAK, lymphokine-activated killer cells were prepared by incubating splenocytes with IL-2 (1,000 U per ml) for 5 days; PD, Abelson-MuLV transformed pre-B cell line; WEHI 3, macrophage/monocyte line; Mac26, a M-CSF dependent bone marrow derived macrophage line; Ba/F3, an IL-3-dependent lymphoid line; FDCP3, a GM-CSF dependent myeloid line; 3T3, a fibroblast line from C3H mouse. *c*, S1-nuclease analysis of RNA from different murine tissues. The structure of the 5' end-labelled probe is shown diagrammatically at the bottom of the figure. The boxed portion indicates *ltk* cDNA and the single line indicates the plasmid sequence. 100 µg total cellular RNA was hybridized at 60 °C with single-stranded <sup>32</sup>P(5' end)-labelled fragment according to Berk and Sharp<sup>33</sup>. S1 digestion was carried out at 15 °C for 90 min with 500 U per ml of enzyme and the protected fragments were analysed by electrophoresis on denaturing polyacrylamide gels. Lanes are (1)  $\phi$ X174 *Hae*III marker; (2) probe only; (3) thymus RNA; (4) LAK RNA; (5) spleen RNA; (6) kidney RNA; (7) liver RNA; (8) yeast transfer RNA control. The protected fragment is marked by an arrow.

protein, but the 64K band may represent either a second *ltk* gene product due to alternative splicing, or a cross-reactive protein.

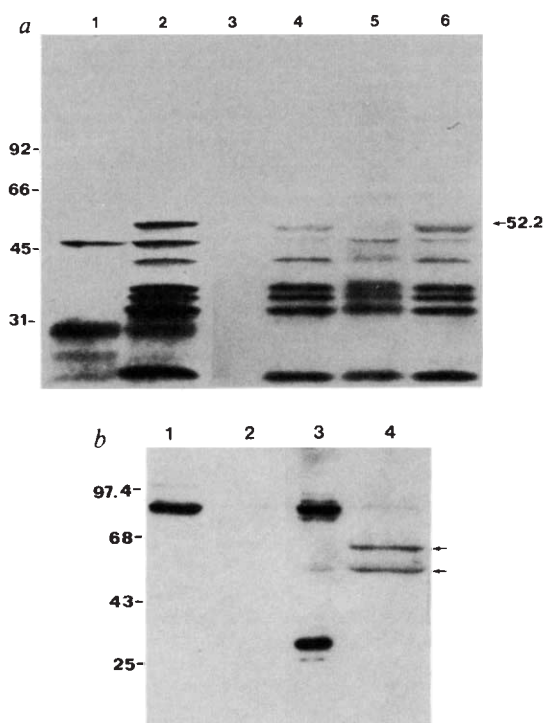
These results show that authentic *ltk* transcripts exist in thymocytes encoding a protein with features of a transmembrane catalytic subunit. Noteworthy in this respect is the analogy of *ltk* to the  $\beta$ -chain of the insulin receptor. The  $\beta$ -chain itself is a transmembrane kinase devoid of a ligand recognition unit, but is linked to the  $\alpha$  chain containing the insulin binding site<sup>2,3</sup>. The  $\alpha$  and  $\beta$  chains of the insulin receptor are processed from a single polypeptide, the product of a single gene. Should *ltk* likewise be linked to another chain, this chain would probably be encoded by a separate gene.

The control of haemopoiesis in the microenvironment of the bone marrow and lymphatic organs involves secreted haematopoietins and possibly cellular interactions between

stromal and haematopoietic precursor cells. Several haematopoietin receptors have a tissue distribution similar to *ltk*. These include the receptors for interleukins 2, 3 and 4 which, like *ltk*, are expressed at variable levels in haematopoietic cells<sup>26-28</sup>. The reported relative molecular mass values relating to the interleukin receptors are between 50K and 80K: 50K and 75K for the IL-3 receptor<sup>27</sup>, 60K for the BSF-1/IL-4 receptor<sup>28</sup>, and 55K and 75K for the  $\alpha$  and  $\beta$  chains of the IL-2 receptor respectively<sup>26</sup>. Furthermore, as the IL-2 and IL-3 receptors comprise two chains<sup>26,27</sup>, one chain could predominate in ligand recognition and the other contribute mainly to signal transduction. It is possible that *ltk* may provide the latter function for one or several of the haematopoietic receptors.

We thank Iris Nefesh for help in screening the library, Drs Eli Kedar, Ruth Gallily and Shlomo Z. Ben-Sasson for cell lines, Drs David W. Weiss, George Klein and Dov Sulitzeanu for





**Fig. 4** The *ltk* protein of the thymus has properties of an integral membrane protein. *a*, Immunoprecipitation of *ltk* protein synthesized *in vitro* in the presence or absence of thymus proteins fractionated in Triton X-114 detergent. Lane 1, *In vitro* translation tRNA control; lane 2, *in vitro* translation of *ltk* RNA; lanes 3–6, immunoprecipitations of *ltk* *in vitro* translation products with pre-immune serum (lane 3); anti-*ltk* serum (lane 4); anti-*ltk* in the presence of the detergent phase proteins (lane 5); anti-*ltk* in the presence of the aqueous phase proteins (lane 6). *b*, Western blot analysis of the thymus *ltk* protein. Lanes 1 and 2, reaction with pre-immune serum; lanes 3 and 4, reaction with anti-*ltk* serum; aqueous phase proteins (lanes 1 and 3); detergent phase proteins (lanes 2 and 4). Sizes of standards are shown to the left (K).

**Methods.** The entire cDNA insert of pRK14 was subcloned into pIBI-30 (in the sense orientation) and the plasmid linearized with *Bam*HI. *In vitro* transcripts were synthesized by T7 RNA polymerase using the Ribogenesis II Biosystem (International Biotechnologies). RNA (200 ng) was used for *in vitro* translation in nuclease-treated rabbit reticulocyte lysate (Stratagene) with [<sup>35</sup>S]methionine (30  $\mu$ Ci, 1,100 Ci mmol<sup>-1</sup>, New England Nuclear). Rabbit antiserum specific to the C-terminal portion of the putative *ltk* protein was prepared, using a  $\beta$ -galactosidase fusion protein containing the last 113 amino acids of the predicted *ltk* protein. Products of *ltk* *in vitro* synthesis were immunoprecipitated either with pre-immune serum (10  $\mu$ l) as a control, or with anti-*ltk* serum (10  $\mu$ l) for 16 h at 4 °C, followed by 2 h incubation with protein A Sepharose at 4 °C. Competition studies were performed with the addition to the immunoprecipitation reaction of either the Triton X-114 detergent phase proteins of a murine whole-thymus extract (100  $\mu$ g protein) or the aqueous phase proteins (1 mg protein) in Triton X-114 solution. Following incubation, the Sepharose beads were washed five times with a solution containing 1% Triton X-100, 0.1% SDS, 0.5 M NaCl, 50 mM HEPES, pH 7.4, 1 mg ml<sup>-1</sup> bovine serum albumin and once with 1% Triton X-100 in PBS, then boiled in SDS-2-mercaptoethanol sample buffer. Samples were analysed on a 7.5–15% SDS-polyacrylamide gel which was fluorographed with Amplify (Amersham) and exposed to Agfa-Curix X-ray film for 12 h. For Western analysis, the thymus proteins (50  $\mu$ g) used in the competition assay were separated on a similar gel, transferred to a nitrocellulose filter and reacted with a 1:50 dilution (in 5% skimmed milk in PBS) of either the pre-immune or the anti-*ltk* serum which had been pre-adsorbed on thymus aqueous phase proteins. Blots were then washed three times with the diluent solution, reacted with <sup>125</sup>I-labelled goat anti-rabbit antibodies, washed three times and exposed to Agfa-Curix X-ray film for 3 h.

advice and encouragement, Dr Yoav Citri for critical reading of the manuscript, and Esther Avigad for typing the manuscript. This work was supported by grants from the Israel Cancer Research Fund, the Israel Cancer Association and the Society of Research Associates of the Lautenberg Center. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries.

Received 8 March; accepted 4 May 1988.

- Hunter, T. *Cell* **50**, 823–829 (1987).
- Ebina, Y. *et al.* *Cell* **40**, 747–758 (1985).
- Ulrich, A. *et al.* *Nature* **313**, 756–761 (1985).
- Ulrich, A. *et al.* *Nature* **309**, 418–425 (1984).
- Yarden, Y. *et al.* *Nature* **323**, 226–232 (1986).
- Ulrich, A. *et al.* *EMBO J.* **5**, 2508–2512 (1986).
- Sherr, C. J. *et al.* *Cell* **41**, 665–676 (1985).
- Snyder, M. A., Bishop, J. M., Colby, W. W. & Levinson, A. D. *Cell* **32**, 891–901 (1983).
- Cross, F. R. & Hanafusa, H. *Cell* **34**, 597–607 (1983).
- Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G. Q. & Baltimore, D. *Cell* **44**, 577–586 (1986).
- Whetton, A. D., Bazill, G. W. & Dexter, M. T. *EMBO J.* **3**, 409–413 (1984).
- Dexter, M. T. *et al.* in *Oncogenes and Growth Control* (eds Kahn, P. & Graf, T.) 163–169 (Springer, Heidelberg, 1986).
- Goldfine, I. D. *Endocrine Rev.* **8**, 235–255 (1987).
- Neckameyer, W. S. & Wang, L. H. *J. Virol.* **53**, 879–884 (1985).
- Hopp, T. P. & Woods, K. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3821–3824 (1981).
- Matsushima, H., Wang, L. H. & Shibuya, M. *Molec. cell. Biol.* **6**, 3000–3004 (1986).
- Hafen, E., Basler, K., Edstroem, J. & Rubin, G. M. *Science* **236**, 55–63 (1987).
- Martin-Zanca, D., Hughes, S. H. & Barbacid, M. *Nature* **319**, 743–748 (1986).
- Yu, K. T., Peters, M. A. & Czech, M. D. *J. biol. Chem.* **261**, 11341–11349 (1986).
- Rosen, O. M., Herrera, R., Olowe, Y., Peiruzelli, L. M. & Cobb, M. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3237–3240 (1983).
- Sibley, D. R., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *Cell* **48**, 913–922 (1987).
- Azar, Y. *et al.* *Cell. Immun.* **65**, 194–200 (1981).
- Schejter, E. D. *et al.* *Cell* **46**, 1091–1101 (1986).
- Bordier, C. *J. biol. Chem.* **256**, 1604–1607 (1981).
- Gething, M. J., McCammon, K. & Sambrook, J. *Cell* **46**, 939–950 (1986).
- Smith, K. A. *Immunol. Today* **8**, 11–13 (1987).
- Nicola, N. A. *Immunol. Today* **8**, 135–140 (1987).
- Paul, W. E. & Ohara, J. A. *Rev. Immun.* **5**, 429–459 (1987).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
- Mizusawa, S., Nishimura, S. & Seela, F. *Nucleic Acids Res.* **14**, 1319–1324 (1986).
- Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767–4771 (1987).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).

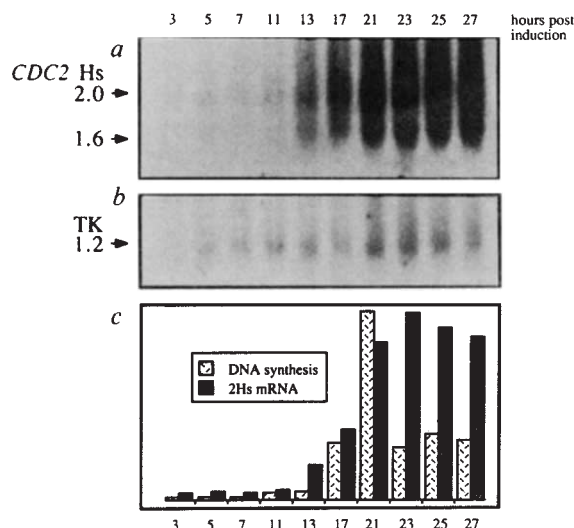
## Regulated expression and phosphorylation of a possible mammalian cell-cycle control protein

Melanie G. Lee\*, Christopher J. Norbury\*,  
Nigel K. Spurr† & Paul Nurse\*

\* ICRF Cell Cycle Control Laboratory, Microbiology Unit,  
Department of Biochemistry, South Parks Road,  
Oxford OX1 3QU, UK

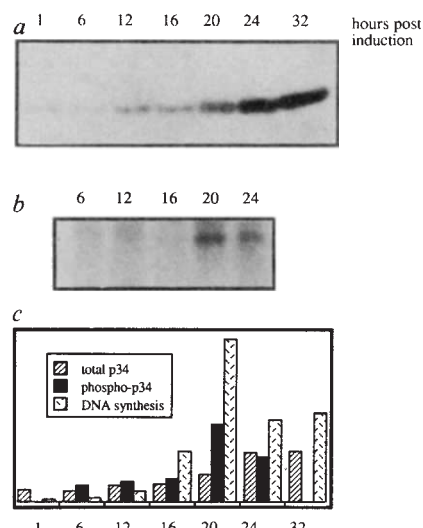
† Imperial Cancer Research Fund, Clare Hall Laboratories,  
South Mimms, Herts EN6 3LD, UK

A novel approach to the study of the control of the mammalian cell cycle was opened by the cloning of a human gene by complementation of a fission-yeast *cdc2* cell-cycle mutant<sup>1</sup>. We have investigated the behaviour of the RNA and protein products of this human gene, *CDC2Hs*, and its murine equivalent, *CDC2Mm* during serum starvation and re-feeding of cultured fibroblasts. In contrast to the pattern of wild-type *cdc2*<sup>+</sup> expression in fission yeast previously described<sup>2,3</sup>, the mammalian homologue displays variation in both RNA and protein levels during exit from and re-entry into the mitotic cycle. Like its yeast counterpart, however, the mammalian *CDC2* protein (p34<sup>CDC2</sup>) becomes dephosphorylated upon shifting from exponential growth to quiescence, and rephosphorylated late in the G<sub>1</sub> phase when cells are stimulated to re-enter the cycle. We propose that phosphorylation of p34<sup>CDC2</sup> serves as a regulatory mechanism generally in eukaryotic cells, while transcriptional control of the *CDC2* gene in higher eukaryotes may be relevant to long term processes such as senescence and differentiation.



**Fig. 1** Induction of *CDC2Hs* transcripts following serum stimulation. Subconfluent ICRF23 human fibroblasts were rendered quiescent by incubation for 48 h in Dulbecco modified Eagle's medium (DMEM) supplemented with 0.5% fetal calf serum (FCS). After addition of FCS to 10%, cells were collected at the times indicated and total cytoplasmic RNA was prepared. 40  $\mu$ l samples from each time-point were subjected to 1% agarose formaldehyde gel electrophoresis and transferred to a nylon filter (Hybond N, Amersham). Autoradiographs shown were obtained after probing the filter with radiolabelled *CDC2Hs* (a) or a human thymidine kinase (TK) cDNA (b). The relative levels of *CDC2Hs* transcripts in each sample were determined by scanning the autoradiograph shown in (a) with an LKB Ultrosan XL densitometer (c). Levels of DNA synthesis, as judged by the incorporation of  $^3$ H-thymidine (1 h pulse at 5  $\mu$ Ci ml $^{-1}$ ) into acid-precipitable material, were also determined for each timepoint, and are displayed in the same column graph. The y-scale for each parameter is linear, with zero at the origin. Densitometric scanning of (b) showed that TK RNA levels rose ~three-fold between 3 and 23 h after serum addition (data not shown).

The ability of plasmids from a human complementary DNA library to complement a cell-cycle mutation in the fission yeast *Schizosaccharomyces pombe* has been used to identify a human homologue of the yeast *cdc2<sup>+</sup>* gene<sup>1</sup>. The latter gene function is required twice in each *Schiz. pombe* mitotic cycle<sup>4</sup>, firstly at a point in *G*<sub>1</sub> called 'start' about 15–30 min before S phase, and again at a second control point in *G*<sub>2</sub>. A homologous gene function *CDC28* has been identified in budding yeast (*Saccharomyces cerevisiae*) and both *cdc2<sup>+</sup>* and *CDC28* have been shown to encode a protein kinase of relative molecular mass (*M*<sub>r</sub>) 34,000 (34K)<sup>3,5</sup>. The human *cdc2<sup>+</sup>* homologue, *CDC2Hs*, has a predicted protein product (p34<sup>*CDC2Hs*</sup>) which display 63% identity with the amino-acid sequence of the yeast p34<sup>*cdc2*</sup>, and an antiserum directed against the perfectly conserved peptide EGVSTAIRESLLKE (referred to here as the PSTAIR antiserum) detects a protein of *M*<sub>r</sub> 34K in extracts from a number of human cell lines<sup>1</sup>. Other antibodies raised against the yeast *cdc2<sup>+</sup>* and *CDC28* proteins also detect a protein of the same size in human cells<sup>6</sup>. The PSTAIR antiserum detects a p34 species in Western blots and immunoprecipitates from murine cells (Fig. 3a, c). Since the free PSTAIR peptide competed specifically with the p34 species detected in these cells for binding to this antiserum (Fig. 3c), it is probable that this species is the authentic product of the murine *cdc2<sup>+</sup>* homologue, which we term *CDC2Mm*. The start control point in yeast may be analogous to previously reported mammalian *G*<sub>1</sub> controls, in particular the restriction (R) point<sup>7</sup>, the pm to ps transition<sup>8</sup> and the W point<sup>9–11</sup>. Beyond these points fibroblasts become committed to initiate a new round of DNA synthesis even if the growth stimuli which caused the initial mitogenic response are



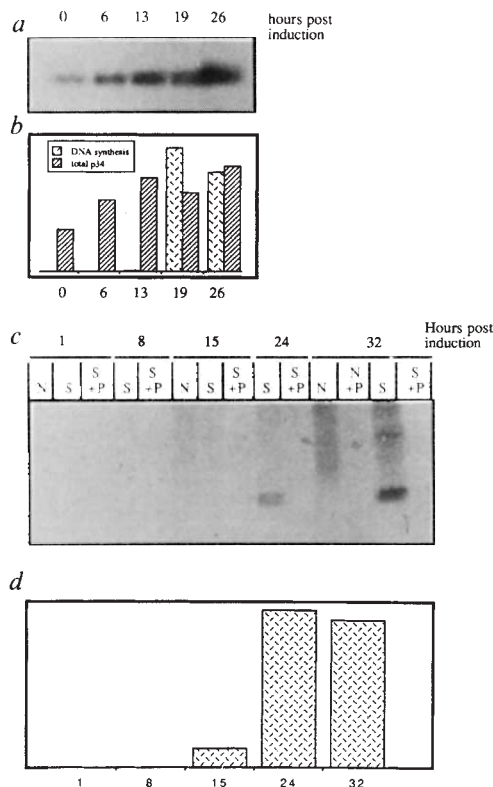
**Fig. 2** Total protein levels and phosphorylation of p34<sup>*CDC2Hs*</sup> after serum stimulation of quiescent ICRF23 fibroblasts. a, Protein extracts from cells treated as in Fig. 1 were separated by 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE), transferred to a nylon membrane (GeneScreen Plus, NEN) and immunoblotted according to the manufacturer's instructions using the PSTAIR antiserum. This single band in each track corresponds to p34<sup>*CDC2Hs*</sup>. b, Cells were prepared in the same way as in Fig. 1, except that during the last 12 h of serum starvation the medium was replaced with phosphate-free DMEM supplemented with 0.5% dialysed FCS and 250  $\mu$ Ci ml $^{-1}$  carrier-free  $^{32}$ P-ortho-phosphate (Amersham). Dialysed FCS was then added to 10%, and cell extracts prepared (as previously described<sup>1</sup>) at the times shown were immunoprecipitated with the PSTAIR antiserum, subjected to 10% SDS-PAGE and autoradiographed. c, Densitometric scanning of (a) and (b) as described in Fig. 1 was used to compare relative levels of total p34<sup>*CDC2Hs*</sup> and phospho-p34<sup>*CDC2Hs*</sup> with DNA synthesis at the times shown.

removed. We have investigated aspects of mammalian *CDC2* expression to determine the extent to which they correlate with cell-cycle progression, in particular through the late-*G*<sub>1</sub> phase.

There does not appear to be any transcriptional control of *cdc2* function in *Schiz. pombe*<sup>2</sup>. In contrast, *CDC2Hs* transcripts were present at barely detectable levels in ICRF23 human fibroblasts<sup>12</sup> which had been in a quiescent (*G*<sub>0</sub>) state for 48 h (Fig. 1a); by 23 hours after serum stimulation, an equivalent amount of total-cell RNA contained 25-fold more *CDC2Hs* transcripts. This peak of transcripts coincided with maximal levels of RNA encoding thymidine kinase, a known late-*G*<sub>1</sub> marker<sup>13</sup>, though the latter increased only threefold over the same time course (Fig. 1b). The pattern of RNA abundance is not, however, directly reflected in total p34<sup>*CDC2Hs*</sup> protein levels, which increased only fivefold under the same induction conditions (Fig. 2). Thus, although p34<sup>*CDC2Hs*</sup> levels decrease in response to serum starvation, the protein is nonetheless readily detectable in quiescent cells, unlike its transcript.

While the mammalian homologue of *cdc2<sup>+</sup>* was isolated from a human cDNA library<sup>1</sup>, murine cells have proved more useful than human cells in studies of the mammalian mitotic cycle. In particular, murine 3T3 cell lines have been widely used and are comparatively well understood in terms of their behaviour during serum stimulation from quiescence<sup>7,14,15</sup>. For this reason we have investigated the expression of *CDC2Mm* in Swiss 3T3 cells<sup>16</sup>. In these cells there was only a twofold variation in total p34<sup>*CDC2Mm*</sup> during serum stimulation (Fig. 3a). The persistence of p34<sup>*CDC2*</sup> in quiescent human and mouse cells (Figs 2a, 3a) is reminiscent of the behaviour of the homologous protein in yeast<sup>3</sup>, and suggests that the functions of p34 in the cell cycle



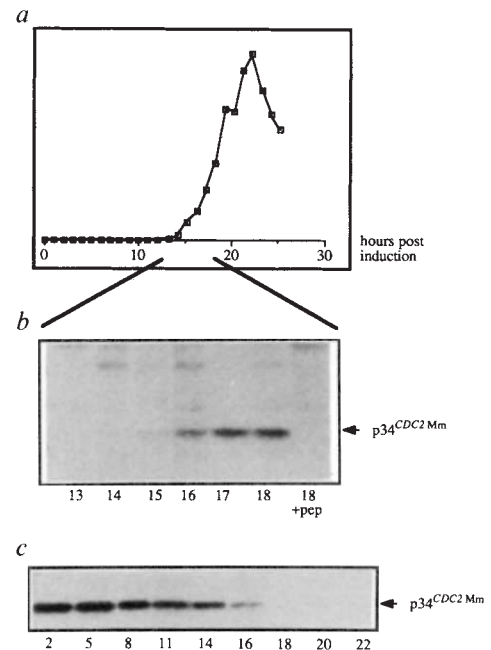


**Fig. 3**  $p34^{CDC2Mm}$  in Swiss 3T3 cells during serum stimulation. *a*, Immunoblotting analysis was performed as described in Fig. 2 after stimulation of quiescent Swiss 3T3 cells with 10% FCS. Densitometric scanning of the autoradiograph (*b*) showed that  $p34^{CDC2Mm}$  levels increased only ~twofold by 26 h of stimulation, and did not correlate with levels of DNA synthesis. *c*, After serum starvation and  $^{32}P$ -phosphate labelling as described in Fig. 2, Swiss 3T3 cell extracts were either stimulated with 10% dialysed FCS (S) or were not stimulated (N). Cell extracts prepared after the times shown were then used in immunoprecipitations with the PSTAIR antiserum or with PSTAIR antiserum pre-incubated with the free PSTAIR peptide (+P). The bands present in the 24 and 32 h serum-stimulated tracks correspond to phosphorylated  $p34^{CDC2Mm}$ . *d*, DNA synthesis assays were performed at the times indicated after serum stimulation; the results are presented in the form of a column graph as in Fig. 1.

are unlikely to be mediated solely by changes in the abundance of the protein in these systems.

The  $p34^{cdc2}$  product of the *cdc2*<sup>+</sup> gene of fission yeast is a protein kinase the activity of which may be regulated by phosphorylation<sup>3</sup>. Specifically,  $p34^{cdc2}$  phosphorylation is seen to decrease during withdrawal from the cell cycle to stationary phase in response to nitrogen starvation. To study the behaviour of mammalian  $p34^{CDC2}$  upon shifting to  $G_0$ ,  $^{32}P$ -phosphate-labelled Swiss 3T3 cells were transferred to low-serum medium, and cell extracts were immunoprecipitated with the PSTAIR antiserum (Fig. 4c). While phosphorylated  $p34^{CDC2Mm}$  was readily detected in cells which remained in exponential growth, the phosphoprotein disappeared between 10 and 22 hours after serum starvation. Note that the  $p34^{CDC2Mm}$  protein was readily detected by immunoblotting after at least 48 hours of serum starvation (Fig. 3a). Thus  $p34^{CDC2}$  dephosphorylation accompanies transition from exponential growth to the  $G_0$  state in both yeast and mammals.

We also determined the timing of  $p34^{CDC2}$  phosphorylation in mammalian cells during serum stimulation from quiescence. Initial experiments showed that phospho- $p34^{CDC2}$  was readily



**Fig. 4** The timing of phosphorylation and dephosphorylation of  $p34^{CDC2Mm}$  in Swiss 3T3 cells. *a*, An accurate profile of DNA synthesis in Swiss 3T3 cells after serum stimulation (at time zero) from quiescence;  $^3H$ -thymidine incorporation is represented on the linear y-axis. *b*, Phosphate labelling and immunoprecipitations were performed on a parallel batch of cells as described in Fig. 3; the phospho- $p34^{CDC2Mm}$  band indicated was detectable from 14 h onwards. Half of the 18 h sample was immunoprecipitated with PSTAIR antiserum which had been pre-incubated with the free PSTAIR peptide (18+pep). *c*, Phosphate-labelled quiescent Swiss 3T3 cells were stimulated with 10% dialysed FCS; after 24 h the medium was replaced with DMEM supplemented with 0.5% dialysed FCS and 250  $\mu Ci ml^{-1}$  carrier-free orthophosphate. At the times shown (h) after the shift back to low-serum medium, cell extracts were immunoprecipitated as described in Fig. 2; phospho- $p34^{CDC2Mm}$  was largely undetectable by 18 h after the shift.

detectable at 20 and 24 h after serum induction in ICRF23 fibroblasts (Fig. 2b) and at 24 and 32 h after serum induction in Swiss 3T3 cells (Fig. 3c), correlating approximately with maximal DNA synthesis. A more precisely defined time course (Fig. 4b) showed that phospho- $p34^{CDC2Mm}$  is first detectable 14 h after serum stimulation of quiescent Swiss 3T3 cells. In this system, S phase begins about 15 h after serum addition, with maximal DNA synthesis being detected at 21 h (Fig. 4a); it appears, then, that phosphorylation of  $p34^{CDC2Mm}$  first occurs just before the  $G_1$  to S transition, at about the same time or just after cells are passing the late- $G_1$  control points<sup>7-11</sup>. Thus in both yeast and mammalian cells, re-entry into the mitotic cycle is accompanied by  $p34$  phosphorylation.

The changes in  $p34^{CDC2}$  we have described are clearly associated with re-entry into the cell cycle, suggesting that this protein is required for progression through the cycle, although our experiments do not allow us to conclude that it is directly required for DNA synthesis. Given its causative role over entry into S phase in the yeasts, however, it is likely that it plays a similar role in mammalian cells. Changes in the phosphorylation state, rather than the absolute abundance of the *CDC2* protein, appear to be of central importance in this control, as has been observed in fission yeast<sup>3</sup>. These changes occur just before the  $G_1$  to S-phase transition, as in the yeasts, which could be correlated with passage through the W point<sup>9-11</sup>, or even the earlier R point<sup>7</sup> or pm to ps transition<sup>8</sup>. On average these earlier

controls occur 2–3 hours before S phase, but this period is very variable and some cells initiate DNA synthesis immediately after passing the pm to ps transition<sup>8</sup>. This timing is consistent with the pattern of increase we have observed in p34<sup>CDC2</sup> phosphorylation. We therefore suggest that when mammalian cells are stimulated to leave quiescence and enter the cell cycle, on passing the late G<sub>1</sub> control points p34<sup>CDC2</sup> becomes activated by phosphorylation. The transcriptional regulation which we have observed may influence p34<sup>CDC2</sup> abundance in the longer term, particularly *in vivo*. For example reduced transcript levels of *CDC2* could serve to direct cells towards quiescence for long periods of time *in vivo*, towards a differentiated state, or towards senescence. We therefore propose two levels of control of p34<sup>CDC2</sup>, one acting in the short term involving phosphorylation of the protein and second acting in the long term involving regulation of transcript levels.

We thank Lesley Rooke for help with the ICRF 23 time course experiments, Mike Fried for his willing provision of incubator

space and Stuart MacNeill for useful comments during the preparation of this manuscript.

Received 29 March; accepted 5 May 1988.

1. Lee, M. G. & Nurse, P. *Nature* **327**, 31–35 (1987).
2. Durkacz, B., Carr, A. & Nurse, P. *EMBO J.* **5**, 369–373 (1986).
3. Simanis, V. & Nurse, P. *Cell* **45**, 261–268 (1986).
4. Nurse, P. & Bissett, Y. *Nature* **292**, 558–560 (1981).
5. Reed, S. I., Hadwiger, J. A. & Lorincz, A. T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4055–4059 (1985).
6. Draetta, G., Brizuela, L., Potashkin, J. & Beach, D. *Cell* **50**, 319–325 (1987).
7. Pardee, A. B., Dunbrow, R., Hamlin, J. L. & Kletzien, R. F. *Rev. Biochem.* **47**, 715–750 (1978).
8. Zetterberg, A. & Larsson, O. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5365–5369 (1985).
9. Pledger, W., Stiles, C., Antoiades, H. & Scher, C. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2839–2843 (1978).
10. Das, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5677–5681 (1981).
11. Campisi, J. & Pardee, A. *Molec. cell. Biol.* **4**, 1807–1814 (1984).
12. Taylor-Papadimitriou, J., Shearer, M., Balkwill, F. R. & Fantes, K. H. *J. Interferon Res.* **2**, 479–491 (1982).
13. Coppock, D. L. & Pardee, A. B. *Molec. cell. Biol.* **7**, 2925–2932 (1987).
14. Holley, R. S. *Nature* **258**, 478–480 (1975).
15. Rozengurt, E. *Curr. Topics cell Regul.* **17**, 59–88 (1980).
16. Todaro, G. J. & Green, H. *Virology* **23**, 117–119 (1964).

## Monomeric insulins obtained by protein engineering and their medical implications

J. Brange, U. Ribel, J. F. Hansen, G. Dodson\*, M. T. Hansen, S. Havelund, S. G. Melberg, F. Norris, K. Norris, L. Snel, A. R. Sørensen & H. O. Voigt

Novo Research Institute, Novo Alle, DK-2880 Bagsvaerd, Denmark

\* Department of Chemistry, University of York, York YO1 5DD, UK

The use of insulin as an injected therapeutic agent for the treatment of diabetes has been one of the outstanding successes of modern medicine. The therapy has, however, had its associated problems, not least because injection of insulin does not lead to normal diurnal concentrations of insulin in the blood. This is especially true at meal times when absorption from subcutaneous tissue is too slow to mimic the normal rapid increments of insulin in the blood. In the neutral solutions used for therapy, insulin is mostly assembled as zinc-containing hexamers<sup>1</sup> and this self-association, which under normal physiological circumstances functions to facilitate proinsulin transport, conversion and intracellular storage<sup>2</sup>, may limit the rate of absorption. We now report that it is possible, by single amino-acid substitutions, to make insulins which are essentially monomeric at pharmaceutical concentrations (0.6 mM) and which have largely preserved their biological activity. These monomeric insulins are absorbed two to three times faster after subcutaneous injection than the present rapid-acting insulins. They are therefore capable of giving diabetic patients a more physiological plasma insulin profile at the time of meal consumption.

Experiments to create a monomeric insulin that did not self-associate were undertaken by analysing the interactions between monomers in insulin dimers and in hexamers. The molecules' assembly properties were then modified by specifically altering, through DNA technology, the amino acids whose side chains were involved or affected by dimer (or hexamer) formation. Preliminary accounts have been published in abstract form<sup>3–7</sup>.

As the strongest subunit interactions are between dimerizing monomers, association is most effectively counteracted by preventing dimer formation. The principal interactions involve B8, B9, B12, B13, B16, B23–B28<sup>1,8</sup>. There is evidence that five of these (B12, B16, B23–B25) are included in the receptor binding region of the hormone<sup>1,8–10</sup>. Thus, the targets chosen for mutation were those amino acid residues involved in dimer formation but located on the periphery of the putative receptor binding

region<sup>9</sup>. Such residues include B9 and B12, whose involvement in receptor binding is still unresolved<sup>11</sup>, and B10, B26, B27 and B28, which are located just outside the receptor-binding region. Our aim was to substitute amino acids that favoured the monomer but did not at the same time destabilize its own three-dimensional structure or interfere with its biological activity.

The main strategy followed was to introduce charge repulsion into the monomer-monomer interface. In most cases, negative charges were inserted at positions where they could oppose pre-existing negatively charged carboxyl groups already found adjacent to the dimer-forming interface (Fig. 1b, c, e). In the case of B12Val and B24Phe, which both make mutual contact in the dimer, we recognized that the direct approach by Jones *et al.*<sup>12</sup>, where such side chains are converted to carboxylic acids, would probably produce an insulin with low potency (see Table 1).

The effect of steric hindrance between monomers was investigated by introducing an extra methyl group at B12(Val → Ile).

Figure 1a illustrates the six selected mutation sites on the monomer. The insulin analogues were prepared by single or double substitutions at six different residues by oligonucleotide directed mutagenesis or by total gene synthesis. In addition, one analogue was deamidated in position A21.

The association behaviour of the different analogues was assessed by osmotic pressure measurements (Table 1) and by circular dichroism (CD) spectroscopy (Fig. 2). The mutations have reduced association in all cases, but there is a clear gradation: B12Glu ≥ B9Asp + B27Glu = B9Asp > B28Asp ≥ B27Glu + A21Asp > B26Glu > B10Asp > B12Ile > B27Glu. The first five mutant insulins are essentially monomeric at concentrations of ~10<sup>-4</sup> M and the first three remain essentially monomeric at 10<sup>-3</sup> M (Fig. 2 and Table 1).

The biological activity of the different analogues *in vitro* and *in vivo* is shown in Table 1. Analogues showing reduced, free fat cell (FFC) activity, such as B9Asp and B12Ile, were in a minority and in only one case, that of B12Glu, was activity very low. Only B10Asp has, in accordance with a recent report<sup>13</sup>, increased FFC activity. All analogues, with the probable exception of B12Glu, have full or nearly full biological activity *in vivo*.

The rate of disappearance from the subcutaneous depot after injection of analogue B9Asp + B27Glu compared with human insulin is illustrated in Fig. 3a. The initial rate of absorption (0–20 min) is approximately three times faster for the analogue than for human insulin and the difference in rate of disappearance during the whole course of absorption is highly significant statistically. The faster absorption of the essentially monomeric analogues is confirmed by measurements of the appearance of



**Table 1** Identity, association, biological activity and bioavailability of insulin analogues

| Insulin                            | Amino-acid analysis<br>$\Delta(X-I)$<br>Arg          | Association state deduced from osmometry |                    | Biological activity (% of insulin) |                  | Subcutaneous absorption                       |               |               |
|------------------------------------|------------------------------------------------------|------------------------------------------|--------------------|------------------------------------|------------------|-----------------------------------------------|---------------|---------------|
|                                    |                                                      | $c = 0.2 \text{ mM}$                     | $c = 1 \text{ mM}$ | FFC                                | MBGA             | $T_{25}$<br>(% of human insulin $\pm$ s.e.m.) | $T_{50}$      | $T_{75}$      |
| DPI                                |                                                      | 1.5                                      |                    |                                    |                  | 61 $\pm$ 7                                    | 76 $\pm$ 7    | 107 $\pm$ 8   |
| Sulphated insulin (bovine)         |                                                      | 1.4                                      | 1.4                |                                    |                  | 75 $\pm$ 16                                   | 82 $\pm$ 13   | 101 $\pm$ 14  |
| B9Ser $\rightarrow$ Asp            | Asp + 0.95<br>Ser - 0.87                             |                                          | 1.1                | 26<br>(25-27)                      | 79<br>(70-89)    | 69 $\pm$ 8                                    | 64 $\pm$ 6    | 75 $\pm$ 6    |
| B10His $\rightarrow$ Asp           | Asp + 0.98<br>His - 0.97                             | 2.1                                      | 2.2                | 207<br>(194-221)                   | 98<br>(87-112)   | 53 $\pm$ 10                                   | 56 $\pm$ 8    | 59 $\pm$ 7    |
| B12Val $\rightarrow$ Ile           | Ile + 0.93<br>Val - 0.90                             | 2.3                                      | 3.3                | 29<br>(27-30)                      | 86<br>(76-97)    | 69 $\pm$ 19                                   | 81 $\pm$ 19   | 88 $\pm$ 18   |
| B12Val $\rightarrow$ Glu + des-B30 | Glu + 1.00<br>Val - 1.07                             |                                          | 1.0                | 0.04<br>(0.038-0.043)              |                  | 64 $\pm$ 13                                   | 59 $\pm$ 8    | 58 $\pm$ 6    |
| B26Tyr $\rightarrow$ Glu           | Glu + 0.84<br>Tyr - 1.05                             | 1.4                                      | 2.0                | 125<br>(121-129)                   | 104<br>(96-112)  | 52 $\pm$ 12                                   | 55 $\pm$ 11   | 63 $\pm$ 20   |
| B27Thr $\rightarrow$ Glu           | Glu + 0.81<br>Thr - 0.94                             | 2.9                                      | 4.0                | 108<br>(104-111)                   | 110<br>(100-121) | 79 $\pm$ 10                                   | 76 $\pm$ 6    | 77 $\pm$ 7    |
| B28Pro $\rightarrow$ Asp           | Asp + 0.99<br>Pro - 1.0                              |                                          | 1.3                | 101<br>(96-105)                    | 104<br>(95-114)  | 49 $\pm$ 8                                    | 53 $\pm$ 7    | 66 $\pm$ 8    |
| B9 + B27                           | Asp + 0.96<br>Glu + 0.94<br>Ser - 0.94<br>Thr - 0.93 | 1.1                                      | 1.1                | 31<br>(31-32)                      | 93<br>(86-101)   | 40 $\pm$ 4                                    | 50 $\pm$ 5    | 60 $\pm$ 4    |
| B27 + A21Asn $\rightarrow$ Asp     | Glu + 0.93<br>Thr - 0.94                             | 1.3                                      | 1.5                | 87<br>(82-92)                      | 61<br>(55-66)    | 50 $\pm$ 13                                   | 54 $\pm$ 12   | 67 $\pm$ 12   |
| Zn-free human insulin              |                                                      | 3.1                                      | 4.4                |                                    |                  | 91 $\pm$ 15                                   | 90 $\pm$ 12   | 87 $\pm$ 12   |
| 2 Zn insulin (human)               |                                                      | $\approx$ 6                              | $\approx$ 6        | $\approx$ 100                      | $\approx$ 100    | $\approx$ 100                                 | $\approx$ 100 | $\approx$ 100 |
| 2 Zn insulin (pork)                |                                                      |                                          |                    |                                    |                  | 125 $\pm$ 28                                  | 129 $\pm$ 20  | 118 $\pm$ 13  |

Results of the amino-acid analysis are given as the difference between analogue (X) and human insulin (I) relative to the Arg content of I; only the amino acids involved in the respective mutations are shown.

Figures for the association state represent the ratio between the osmotic pressure of the analogue and six times that of 2 Zn human insulin (hexameric at both concentrations with a calculated mean relative molecular mass of 37,500). Biological activity, relative to human insulin, is shown with 95% confidence limits in brackets.  $T_{25}$ ,  $T_{50}$  and  $T_{75}$  designate the time compared to that of human 2Zn insulin until 25, 50 and 75%, respectively, of the doses injected into pigs disappeared from the subcutaneous depot. Standard errors are estimated from the individual disappearance curves.

Plasmids for expression of single-chain insulin analogue precursors<sup>16,17</sup> were transformed into *S. cerevisiae* strains which were grown in continuous fermentation. The exported precursors were isolated by adsorption to an ion exchange column followed by desorption, precipitation by zinc ions, anion exchange chromatography, transpeptidation to insulin analogue-(B30)-ester<sup>17,18</sup> and reverse phase high performance chromatography (RP-HPLC). The analogues were obtained after cleavage of the ester bond with trifluoroacetic acid, lyophilization, purification on QAE-Sephadex A25 and desalting on Bio-Gel P-4 in 0.5 M acetic acid. One analogue (B27GluA21Asp) was produced by acid hydrolysis (specific deamidation of residue A21) of the analogue precursor B27Glu (1% solution, pH 2.5, 5 weeks at 37 °C) followed by isolation and purification of the hydrolysis product. Purity of the mutants was at least 98% as assessed by RP-HPLC. Identity was verified by a combination of DNA sequencing of the gene<sup>19</sup> and amino-acid analysis (this table). The zinc content of mutants and zinc-free insulin was <0.01 mol per mol insulin; 2 Zn insulin (0.40 mol Zn per mol insulin) was from Novo Industri, Bagsvaerd; despentapeptide (B26-B30) insulin (DPI) was made by Dr S. Tolley, York University, and sulphated insulin was obtained from Connaught-Novo Lab., Toronto.

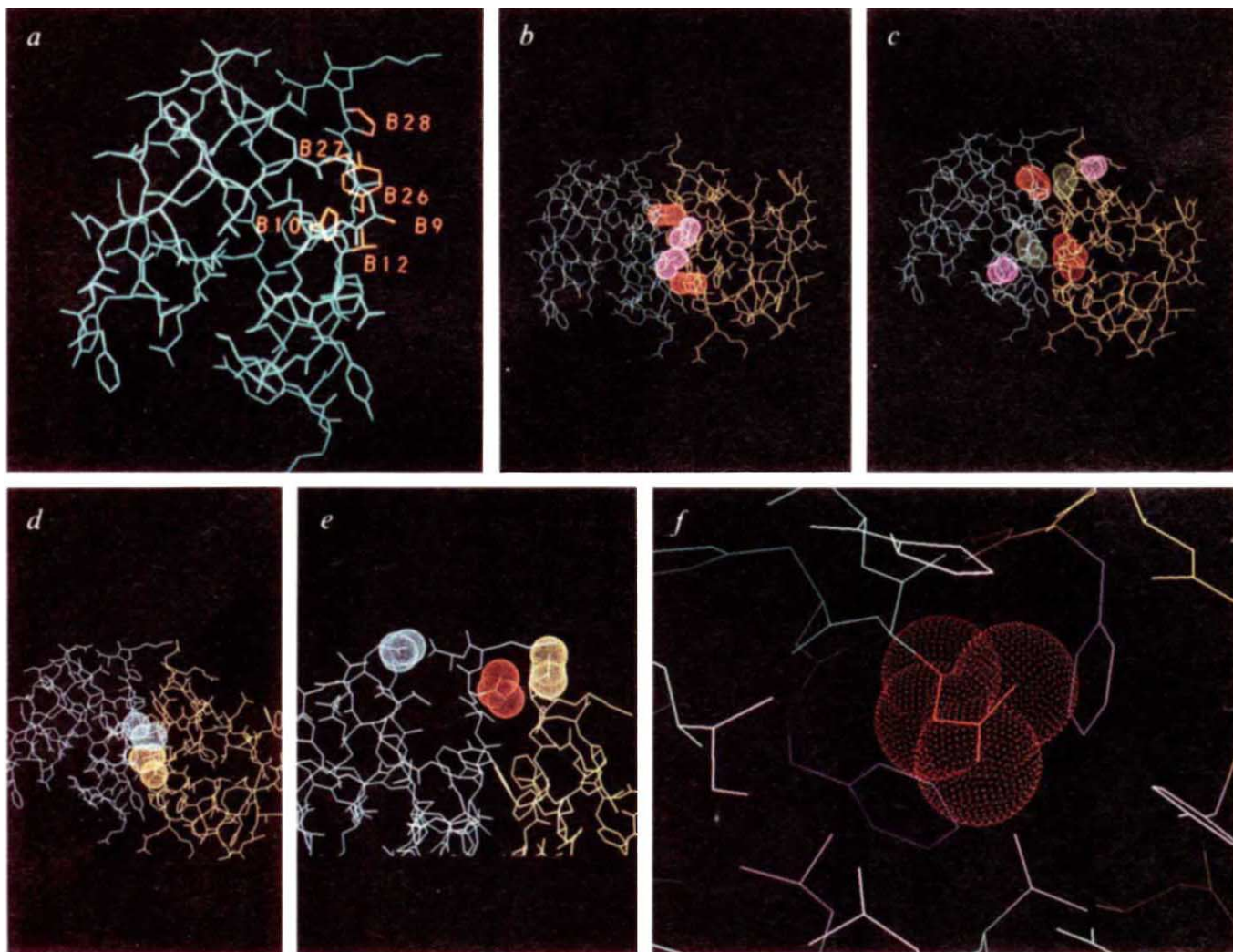
Amino acid analysis (hydrolysis in 6 N HCl for 24 h) was carried out on a Bechmann 121 MB. Osmotic pressure was estimated as mean of at least two measurements on Knauer type 01 osmometer (sample volume 2.5 ml, Schleicher & Schüll membrane RC52, 21 °C) or a single measurement (B12Glu and B26 analogues) on Knauer type 731 (sample volume 0.25 ml, Schleicher & Schüll membrane RC51, 33 °C). Readings were extrapolated to time zero in case of permeation of solute through the membrane<sup>20</sup>.

Biological activity was determined with human insulin as standard by a free fat cell assay (FFC, ref. 21) using mouse adipocytes and a mouse blood glucose assay (MBGA) according to European Pharmacopeia 1984, Assay of insulin, method C. The disappearance of insulin after subcutaneous injection into pigs relative to that of human insulin was measured in a cross over study (see legend to Fig. 3a).

insulin in serum and illustrated for the B28Asp analogue in Fig. 3b. The analogue reaches higher serum values than human insulin, peaks around 30 min, and declines to baseline levels more rapidly than human insulin, which reaches a plateau with its middle point at about 70 min. The corresponding blood glucose levels are shown in Fig. 3c. The times of 25, 50 and 75% absorption of the analogues and two well-known monomeric insulin derivatives after subcutaneous injection relative to that of human insulin are given in Table 1.

Firstly, these results show that charge repulsion by side chains on the periphery of the dimer-forming surface, where the side chains do not make mutual contacts, is an effective mechanism, for reducing association. The closer the repelling groups are

packed in the dimer the more effectively is association reduced (Fig. 1). For example, the most pronounced reduction in dimer formation associated purely with charge is in the B9Asp mutant. This is explained by the dimer packing which brings two B9 + two B13(Glu) residues to within 4 Å (Fig. 1b) and by the hexamer packing which brings six B9 + six B13 side chains closely together in its central core. The larger separation (~10 Å) of B27 from the negatively charged A21 C-terminal group results in the B27Glu mutation having least influence on association (Table 1, Fig. 1c). But the amide group of the A21 side chain is closer to residue B27 (Fig. 1c), and, when A21 is deamidated, Glu substituted in position B27 markedly increases dissociation (Fig. 2b, c; Table 1). The introduction of B28Asp has a more



**Fig. 1** Computer graphics using the FRODO<sup>22</sup> and the HYDRA program<sup>23</sup>. *a*, The human insulin monomer with the side chains of mutated residues shown in red. *b*, Insulin dimer with B13Glu (purple) and mutated B9Asp (red) superimposed in dots indicating the van der Waals surface of the carboxyl groups atoms. Distances between the four individual carboxyl groups are  $\sim 4$  Å after maximizing separation by manual adjustment of the side chains. *c*, Insulin dimer with the A21 residues (C-terminal carboxyl group purple, side chain amide group green) and mutated B27Glu (red) shown as described in (*b*). Distances between the B27 carboxyl group and the A21 C-terminal and A21 amide are 4–6 Å and 9–11 Å respectively. *d*, Insulin dimer with mutated B12Ile (C-gamma and C-delta atoms shown with van der Waals surface). The radius of the C-delta atoms overlaps across dimer interface. *e*, Part of the insulin dimer with mutation B28Asp (red) and nearby negatively charged side chains (A4 blue and B21 yellow) shown as described in (*b*). The distance in the insulin dimer between side chain oxygens in B28 and B21 across interface is  $\sim 3$  Å. The C-terminal B-chain, including B28, has sufficient space to move away from the monomer-monomer interface, but in the case of a B28 negatively charged side chain it will be repelled by A4Glu in its own monomer. *f*, Part of the insulin dimer with mutation B26Glu shown in red as described in (*b*). Aliphatic and aromatic side chains surrounding the mutated and manually fitted residue are shown in purple. The closest carboxylic acids in the dimer are more than 10 Å distance (not shown).

dramatic effect than substitution at B27 (Table 1, Fig. 2*f*). This can be attributed to the closer approach made by B28 (3–4 Å) to the B21Glu of the adjacent monomer in the dimer (Fig. 1*e*).

Secondly, it appears that burial of hydrophilic groups such as glutamic acid at B26 (Fig. 1*f*), whose mutual separation and distance to B21Glu is 10 Å destabilizes the dimer to some extent (Table 1).

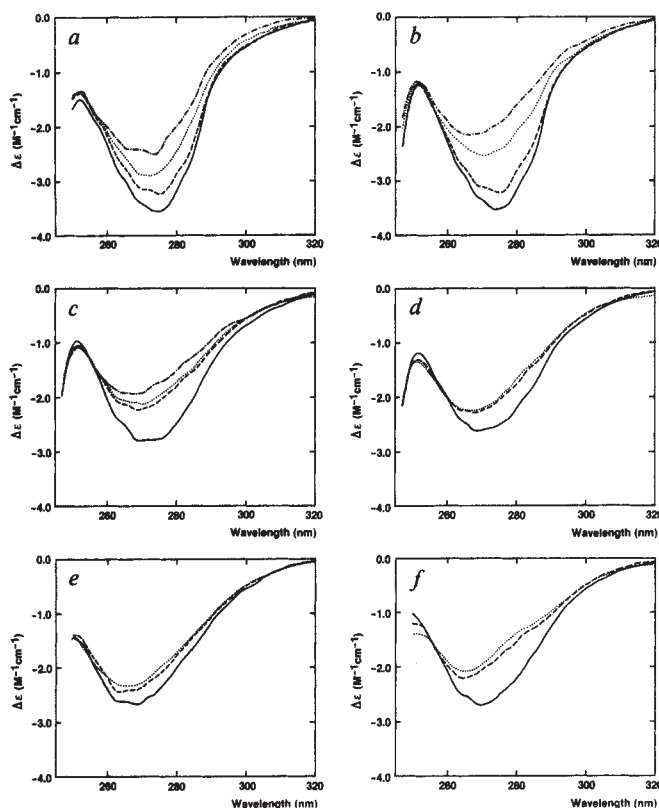
Thirdly, limited steric hindrance associated with substitution of Ile for Val (Fig. 1*d*) is not very effective in reducing dimer formation (Table 1); indeed the B12Ile mutant has been crystallized as a hexamer<sup>14</sup>.

Fourthly, most of these substitutions do not significantly affect the *in vivo* potency of the hormones (Table 1).

Finally and very importantly our finding that insulin molecules with less tendency to self-associate are absorbed substantially faster from subcutis demonstrates for the first time that the size of the insulin unit plays an important role for the rate of transfer of insulin to the blood.

With this discovery the therapeutic goal of achieving rapid increments of plasma-free insulin at the time of meal consumption<sup>15</sup> has been achieved, and has already led to significantly improved glucose control in diabetics (D. R. Owens, personal communication). Thus the monomeric insulins described here are potential candidates for quicker delivery of insulin and better treatment of diabetic patients. Investigation of other properties such as immunogenicity, growth factor and other possible side



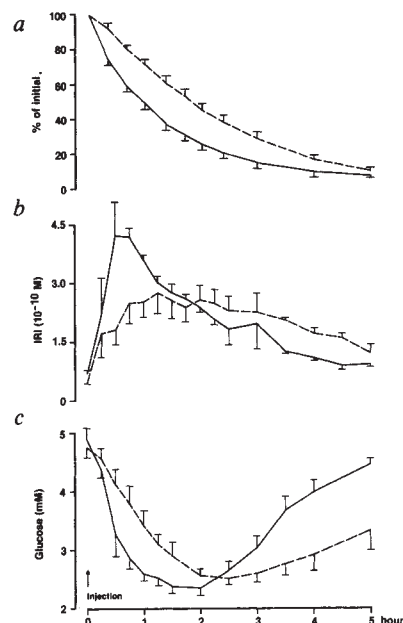


**Fig. 2** Circular dichroism (CD) spectroscopy in the near ultraviolet region. Dissociation of insulin oligomers into monomers as a function of concentration is reflected by the weakening of the negative band at 275 nm (ref. 24) arising from the tyrosines (B16 and B26) in the monomer-monomer interface<sup>25</sup>. *a*, Human insulin: ---,  $3.0 \times 10^{-6}$  M; ···,  $3.0 \times 10^{-5}$  M; ---,  $3.0 \times 10^{-4}$  M; —,  $3.0 \times 10^{-3}$  M. *b*, Mutant B27Glu: ---,  $3.2 \times 10^{-6}$  M; ···,  $3.2 \times 10^{-5}$  M; ---,  $3.2 \times 10^{-4}$  M; —,  $3.2 \times 10^{-3}$  M. *c*, Mutant B27Glu + A21Asp: ---,  $3.0 \times 10^{-6}$  M; ···,  $3.2 \times 10^{-5}$  M; ---,  $3.0 \times 10^{-4}$  M; —,  $3.0 \times 10^{-3}$  M. *d*, Mutant B9Asp: ···,  $3.2 \times 10^{-5}$  M; ---,  $3.2 \times 10^{-4}$  M; —,  $3.2 \times 10^{-3}$  M. *e*, Mutant B9Asp + B27Glu: ···,  $2.7 \times 10^{-5}$  M; ---,  $2.7 \times 10^{-4}$  M; —,  $2.6 \times 10^{-3}$  M. *f*, Mutant B28Asp: ···,  $2.7 \times 10^{-5}$  M; ---,  $2.7 \times 10^{-4}$  M; —,  $2.7 \times 10^{-3}$  M. In *d-f* a 10-fold dilution of the  $10^{-5}$  M solutions did not introduce further spectral changes.

**Methods.** The zinc-free insulins were dissolved in buffer at pH 7.5. Two different buffers were used, 0.050 M NaCl+0.01 M Tris-HCl and 0.025 M Tris-HCl. The spectral changes were independent of the buffer chosen. The concentration was determined photometrically at 276 nm using an extinction coefficient of  $6.2 \times 10^3$  M<sup>-1</sup> cm<sup>-1</sup>. The CD spectra were recorded on a Jobin-Yvon Dichrograph Mark V calibrated with a 0.1% aqueous solution of *d*-10-camphorsulphonic acid ( $\Delta\epsilon(290 \text{ nm}) = 2.36 \text{ M}^{-1} \text{ cm}^{-1}$ ). Spectral band width 1.0 or 2.0 nm, sample temperature 25 °C and cell path length between 10.000 and 0.05 cm. To improve the signal-to-noise ratios spectral data were accumulated and averaged from 5 to 10 scans.

effects is, however, necessary before the utility of these new insulins can be fully established.

We thank M. Christensen for the amino-acid analyses. I. Diers for the yeast fermentations, R. E. Hubbard and L. Nørskov-Lauritsen, for advice on the graphics, I. Jansson and K. Larsen for help with preparing the manuscript. We also thank Y. Broby Madsen, S. Danielsen, B. Draeby Spon, B. Draeger, H. P. Flygnering, L. Gotfred, L. Grønlund Andersen, H. Jensen-Holm, A.-M. Kolstrup, S. Rasmussen, B. Roed, A.-M. Skov Sørensen, A. Stougaard, and I. Wang Hansen for technical assistance and Professor Dorothy Hodgkin for her support and interest in this research.



**Fig. 3** *a*, Absorption rate after subcutaneous injection into pigs (mean  $\pm$  s.e.m.,  $n = 8$ ). *b*, Plasma insulin after subcutaneous injection of B28Asp mutant (—) and human insulin (---) (13 nmol per kg) into fasted pigs (mean  $\pm$  s.e.m.,  $n = 6$ ). *c*, Plasma glucose after subcutaneous injection of B28Asp mutant (—) and human insulin (---) (6.5 nmol per kg) into fasted pigs (mean  $\pm$  s.e.m.,  $n = 6$ ).

**Methods.** Disappearance from subcutis by local external gamma counting<sup>26</sup> of the residual radioactivity of injected <sup>125</sup>I-labelled insulin solutions (0.25 mM  $\sim$  40 i.u. insulin per ml, 0.2% phenol, 1.6% glycerol, pH 7.4) in a pig model<sup>27</sup>. Insulins were labelled with <sup>125</sup>I using the lactoperoxidase method<sup>28</sup>. A dose of 0.7 nmol kg<sup>-1</sup> (corresponding to 0.1 i.u. human insulin per kg) of analogue B9Asp+B27Glu (—) was injected subcutaneously into one side, and human insulin (---) into the other side of the neck of normal pigs and the residual amount followed for 5 h after administration. Immunoreactive insulins (IRI) were determined<sup>29</sup> using individual standard curves of the respective insulins. Glucose was determined by the hexokinase method.

Received 11 January; accepted 27 April 1988.

- Blundell, T. *et al.* *Adv. Protein Chem.* **26**, 279-402 (1972).
- Emdin, S. O. *et al.* *Diabetologia* **19**, 174-182 (1980).
- Brange, J. *et al.* *Protein Engng.* **1**, 238 (1987).
- Brange, J. *et al.* *Diabetes* **26**, Suppl. 1, 77A (1987).
- Hansen, J. F. & Brange, J. *Protein Engng.* **1**, 250 (1987).
- Melberg, S. G. *Protein Engng.* **1**, 255-256 (1987).
- Brange, J. *et al.* *Diabetologia* **30**, 503A (1987).
- Baker, E. N. *et al.* *Phil. Trans. R. Soc.* (in the press).
- Gammeltoft, S. *Physiol. Rev.* **64**, 1321-1378 (1984).
- Fischer, W. H. *et al.* *Biol. Chem. Hoppe-Seyler* **366**, 521-525 (1985).
- Horuk, R. *et al.* *Horon. Cell Reg.* **4**, 123-139 (1980).
- Jones, D. H. *et al.* *Biochemistry* **24**, 5852-5857 (1985).
- Schwartz, G. P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 6408-6411 (1987).
- Derewenda, U. *et al.* *Protein Engng.* **1**, 238-239 (1987).
- Skyler, J. S. *Diabetes Care* **9**, 666-667 (1986).
- Thim, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 6766-6770 (1986).
- Markussen, J. *et al.* in *Peptides 1986*. Proceedings of the Nineteenth European Peptide Symposium (ed. Theodoropoulos, D.) 189-194 (Walter de Gruyter, Berlin, 1987).
- Markussen, J. *et al.* in *Hormone Drugs* (eds Gueriguian, J. L., Bransome, E. D. & Outochoorn, A. S.) 116-126 (United States Pharmacopoeial Convention, Rockville, Maryland, 1982).
- Ribel, U. *et al.* in *Met. Enzym.* **65**, 499-560 (1980).
- Elias, H.-G. in *Characterization of Macromolecular Structure* (ed. F. McIntyre) 28-50 (National Academy of Sciences, Washington, DC, 1968).
- Moody, A. J. *et al.* *Hormone Metab. Res.* **6**, 12-16 (1974).
- Jones, T. A. *J. appl. Crystalllogr.* **11**, 268-272 (1978).
- Hubbard, R. E. in *Computer Aided Molecular Design* (Proceedings of a 2-day conference held in London October 1984) 99-106 (Oyez, London, 1985).
- Wood, S. P. *et al.* *Eur. J. Biochem.* **55**, 531-542 (1975).
- Strickland, E. H. & Mercola, D. A. *Biochemistry* **15**, 3875-3884 (1976).
- Binder, C. *Acta pharmac. tox.* **27**, (Suppl. 2) 1-87 (1969).
- Ribel, U. *et al.* in *Diabetes 1985* (eds Serrano-Rios M. & Lefebvre, P. J.) 891-896 (Elsevier, Amsterdam, 1986).
- Jørgensen, K. H. & Larsen, U. D. *Diabetologia* **19**, 546-554 (1980).
- Heding, L. G. *Diabetologia* **8**, 260-266 (1972).

# Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor

Lukasz Lebioda & Boguslaw Stec

Department of Chemistry, University of South Carolina, Columbia, South Carolina 29208, USA

Enolase or 2-phospho-D-glycerate hydrolase catalyses the dehydration of 2-phosphoglycerate to phosphoenolpyruvate, which in turn is converted by pyruvate kinase to pyruvate. We describe here the crystallographic determination of the structure of yeast enolase at high resolution (2.25 Å) and an analysis of the structural homology between enolase, pyruvate kinase and triose phosphate isomerase. Each of the two subunits of enolase forms two distinctive domains. The larger domain (residues 143–420) is a regular 8-fold  $\beta/\alpha$ -barrel, as first found in triose phosphate isomerase, and later in pyruvate kinase and 11 other functionally different enzymes. An analysis of the molecular geometries of enolase and pyruvate kinase based on the roughly 8-fold symmetry of the barrel showed a structural homology better than expected for proteins related by convergent evolution. We argue that enolase and pyruvate kinase have evolved from a common ancestral multifunctional enzyme which could process phosphoenolpyruvate in both directions along the glycolytic pathway. There is structural and sequence evidence that muconate lactonizing enzyme later evolved from enolase.

Enolase was crystallized as described previously<sup>1</sup>. Native data of 135,350 measured and 21,671 symmetry-independent reflections ( $R_{\text{merge}} = 5.9\%$ ) were collected using a multiwire area diffractometer with three detectors to 2.25 Å resolution. An electron-density map was calculated using isomorphous differences from three heavy-atom derivatives:  $\text{Sm}^{3+}$  (3-Å resolution),  $[\text{PtCl}_4]^{2-}$  (3.5-Å resolution) and Hg (methylmercury to 6-Å resolution), with an overall figure of merit of 0.63. The map was modified by iterative solvent flattening<sup>2</sup>. Owing to the poor quality of the heavy atom derivatives, the map was not fully interpretable, although the molecular outline and secondary structure elements were evident. Subsequent electron-density maps, derived from many cycles of model building, phase combination and restrained least-squares refinement with data

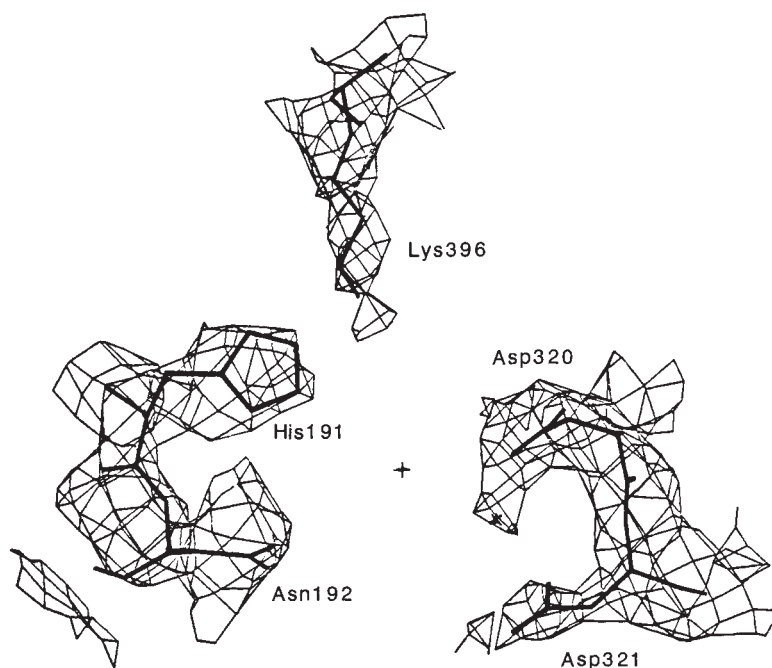
extension to 2.25 Å resolution, permitted the complete structure to be determined and aligned with the known sequence<sup>3,4</sup>, with the exception of three outside loops 35–60, 194–221 and 249–263, which may be partially disordered. An electron-density map of the active site is shown in Fig. 1. At present a standard  $R$ -factor at 2.25 Å resolution is 33% with good stereochemistry. Based on the comparison of different series of least-squares refinement, we estimate that the alpha-carbon positions ( $C_\alpha$ ) of the secondary structure elements are determined with an error of less than 0.3 Å and of the ordered coil segment with less than 0.5 Å error. The coordinates of  $C_\alpha$  have been deposited with the Protein Data Bank (data set 1ENL)) from which copies are available.

The space group of enolase crystals requires the dimeric molecule to be positioned on a crystallographic 2-fold axis. Extensive hydrophobic contacts leave little doubt about the choice of interfaces of the dimer present in solution. The smaller N-terminal domain, built of residues 1–142, starts with a three-stranded antiparallel  $\beta$ -meander. A long  $\alpha$ -helix, broken in the middle by Pro74, and three shorter helices on the inner side of the  $\beta$ -sheet follow. The shorter helices form extensive contacts with the  $\alpha$ -helices of the 8-fold  $\beta/\alpha$ -barrel domain. The C-terminal tail is a relatively short stretch of residues, 421–436, with no secondary structure, packed against the junction of the N-terminal and barrel domains. The overall architecture of the molecule is shown in Fig. 2.

Under physiological conditions, each subunit of enolase has a  $\text{Mg}^{2+}$  cation that is strongly bound in the conformational site and is necessary for catalysis<sup>5</sup>. This cation was exchanged for  $\text{Sm}^{3+}$  in one of the heavy-atom derivatives; its binding site is at the carboxylic end of the  $\beta$ -barrel. The active-site location was further confirmed by crystal-binding experiments with the substrate 2-phosphoglycerate. A difference Fourier map, calculated at 4-Å resolution, had the highest peak 3.0 Å from the  $\text{Mg}^{2+}$  cation. Asp246, Glu295 and Asp320 were identified as  $\text{Mg}^{2+}$  ligands. His191, which is in the active site, is probably the catalytic base<sup>6</sup>. Other residues in the active site region are Asn192, Asp321, Lys345, Arg374 and Lys396.

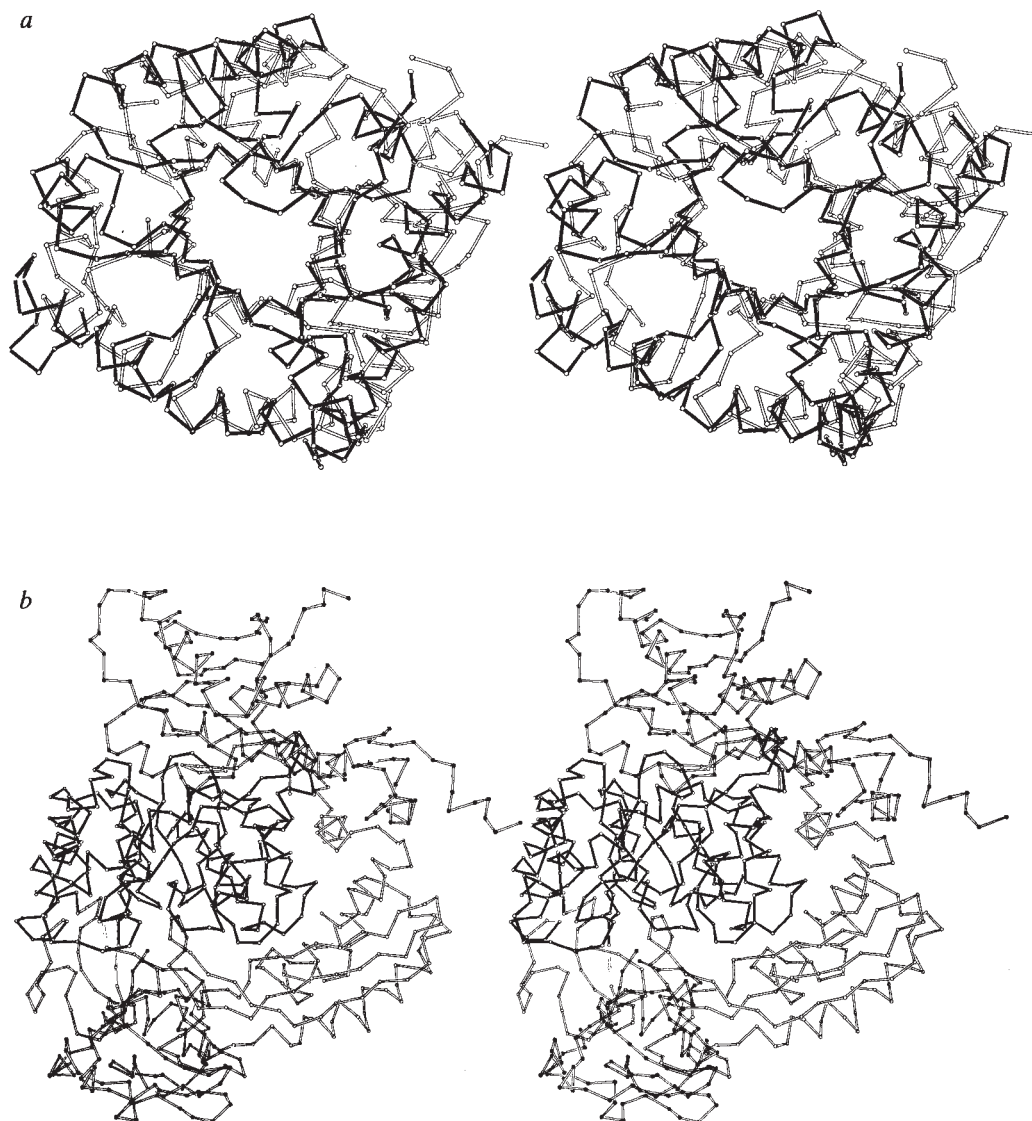
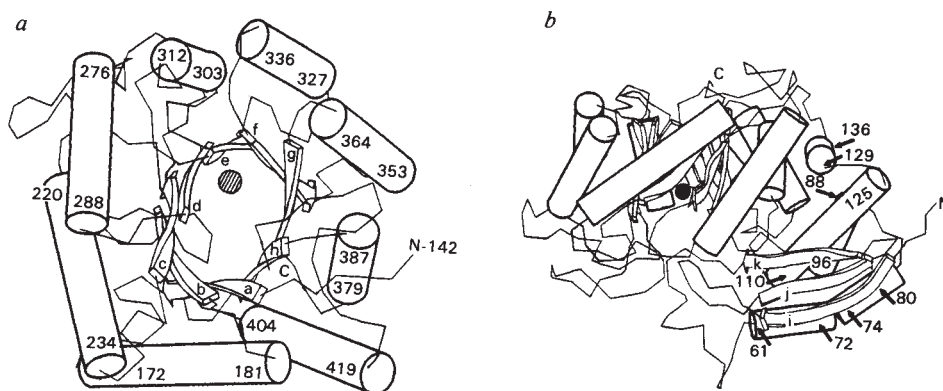
Domains with the 8-fold  $\beta/\alpha$ -barrel fold have been found in 13 functionally different enzymes: triose phosphate isomerase<sup>7</sup>, pyruvate kinase<sup>8</sup>, enolase, aldolase<sup>9,10</sup>, Taka-amylase<sup>11</sup>, xylose isomerase<sup>12</sup>, glycolate oxidase<sup>13</sup>, RuBisCo<sup>14</sup>, muconate lactonizing enzyme<sup>15</sup>, tryptophan synthase<sup>16</sup>, trimethylamine dehy-

**Fig. 1** The active site region. The electron-density map was calculated at 2.8-Å resolution with  $2F_o - F_c$  coefficients and calculated phases. All the atoms displayed were omitted from the model. The substrate binding position is marked with +.





**Fig. 2** The structure of enolase. *a*, 8-fold- $\beta/\alpha$ -barrel domain viewed along the barrel axis shows that symmetry is only approximate. The inner  $\beta$ -barrel is elliptical with the dimensions  $11 \times 8$  Å. *b*, View from a perpendicular direction with both domains included. The sequence on the  $\beta$ -strands is: a, 148-153; b, 189-193; c, 244-248; d, 292-296; e, 316-320; f, 341-345; g, 368-373; h, 394-397; i, 4-11; j, 17-24; k, 29-35.



**Fig. 3** The best superposition of  $C_\alpha$  models of enolase and PK, first  $\beta$ -strand on first, second on second, and so on. *a*, Stereoview of the barrel domains only. Enolase in heavy lines, PK in light. The elliptical  $\beta$ -barrels show an excellent match. *b*, Barrel domain of enolase (in heavy lines) and the other domains of enolase (light lines and light atoms) and PK (light lines and heavy atoms) are included and show the growth of the enzymes in the process of divergent evolution. The PK barrel is omitted for clarity. View along a direction perpendicular to the barrel axis.

Table 1 Comparisons of the structures of the  $\beta/\alpha$ -barrel domains

| Superposed strands | 1-1     | 1-2     | 1-3     | 1-4    | 1-5     | 1-6     | 1-7     | 1-8     |
|--------------------|---------|---------|---------|--------|---------|---------|---------|---------|
| MLE-enolase        | 175 2.4 | 143 3.1 | 127 2.8 | 56 2.9 | 65 2.9  | 53 2.8  | 98 3.2  | 106 3.2 |
| PK-enolase         | 178 3.4 | 152 3.4 | 114 2.9 | 46 2.8 | 85 2.5  | 75 2.9  | 47 2.9  | 116 3.5 |
| TIM-enolase        | 92 3.2  | 140 3.6 | 113 3.2 | 95 2.6 | 100 2.9 | 100 3.1 | 130 3.5 | 118 2.9 |
| TIM-PK             | 159 3.3 | 153 3.2 | 141 3.3 | 67 3.4 | 71 3.1  | 101 3.2 | 130 3.1 | 138 3.2 |

Entries represent the numbers of geometrically equivalent residues with corresponding r.m.s. distances ( $\text{\AA}$ ) between the equivalent  $C_\alpha$  positions below. The barrel domains of enolase (281 residues), PK (247 residues), TIM (287 residues) and MLE (201 residues) were compared in the eight possible superpositions. The enolase barrel has two outside loops, each about 25 residues long, which do not form equivalencies. The MLE has only seven helices. The fit was obtained initially with an interactive computer graphics system and later optimized using the method and criteria described by Rossmann and Argos<sup>29</sup> with fixed parameters  $E1 = 2.0$  and  $E2 = 1.0$ .

drogenase<sup>17</sup>, flavocytochrome  $b_2$ <sup>18</sup>, and N-(5'-phosphoribosyl) anthranilate isomerase-indole-3-glycerol-phosphate synthase<sup>19</sup>. While these impressive data were accumulating, two alternative theories were discussed<sup>8,15,20-23</sup>. The first theory is that of divergent evolution from a common ancestral protein with an 8-fold  $\beta/\alpha$ -barrel fold; the second is convergent evolution to the stable symmetric structure of the barrel convenient for many catalytic functions.

In the process of protein evolution it is well known that the conformation of the backbone is conserved much longer than the amino-acid sequence<sup>24</sup>. As the 8-fold  $\beta/\alpha$ -barrel has pseudo-8-fold symmetry, the superpositions of the structures can be produced by  $45^\circ$  rotation about the 8-fold axis. If two proteins are related by divergent evolution from a common ancestor which deviated from perfect 8-fold symmetry, then the superposition of first strand on first strand, second strand on second strand and so on should be better than the other seven superpositions (such as first strand on second strand and so forth). Such comparisons carried out before for 2-keto-3-deoxy-6-phosphogluconate-aldolase with triose phosphate isomerase (TIM) and pyruvate kinase (PK) (ref. 21) and for muconate lactonizing enzyme (MLE) with TIM (ref. 15) did not show superiority for the natural, or first to first, superposition. Here we report comparisons of structural homology between the glycolytic enzymes enolase, TIM and PK.

To estimate the reliability of this method, we have also compared enolase with MLE as they both have almost the same folding of both domains; the only major difference is that the last  $\alpha$ -helix in the barrel domain is replaced by a coil segment in MLE. When both domains are superimposed, they give 283 equivalent  $C_\alpha$  positions with a r.m.s. deviation of 3.11  $\text{\AA}$ . There is a 15% sequence similarity between enolase and MLE, and the characteristic signature sequence (Ala73-Pro74-Ala75) in enolase, which forces bending of the first  $\alpha$ -helix, is preserved in MLE. MLE is an enzyme involved in degradation of catechol to  $\beta$ -keto adipate<sup>25</sup>; it is very probable that when *Pseudomonas putida* acquired the ability to process catechol, MLE evolved from enolase.

Comparison of the barrel structures listed in Table 1 shows the obvious superiority of a 1-1 superposition of enolase with PK and of enolase with MLE. Moreover, for those two sets of comparisons there is a close similarity of the number of equivalencies as the function of the superposition angle. Somewhat weaker similarity is also observed for TIM and PK. The sequence and structural relatedness suggested a common ancestor for enolase and MLE; analogy of the data in Table 1 for enolase and MLE and for enolase and PK are a strong argument that enolase and PK also evolved from a common ancestor. The best superposition of  $C_\alpha$  models of enolase and PK is shown in Fig. 3. Similarities between enolase and PK are not limited to the folding of the barrel domains. Both enzymes bind the same phosphoenolpyruvate ligand and use two  $Mg^{2+}$  cations per active site (PK also requires  $K^+$ ). The phosphoenolpyruvate binding sites in PK and enolase are similarly located at the carboxylic end of the barrel on strands 2-6. In the best geometrical superposition, the active site Asp112 of PK corre-

sponds to Asp246 of enolase. Another argument which supports origin by divergent evolution and applies to all thirteen enzymes is that the first secondary structure element in the barrel is invariably a  $\beta$ -strand. There is no evidence that a barrel starting with an  $\alpha$ -helix would be handicapped with respect to those starting with a  $\beta$ -strand. Also, almost all enzymes with the  $\beta/\alpha$ -barrel structure activate a hydrogen atom next to a carbonyl group<sup>26</sup>. Although at this point we cannot exclude the possibility that some of the proteins with the  $\beta/\alpha$ -barrel structure evolved independently, we have presented evidence that some of them developed by divergent evolution. There are more similarities between enolase and PK than any other pair of glycolytic enzymes and so it appears that the specialization of those two enzymes was the last one in the development of the glycolytic pathway. Either there existed a multifunctional 'enokin', or a spectrum of enzymes produced from the single gene by an imprecise translation mechanism had both enolase and pyruvate kinase activity.

Evolutionary relationship between glycolytic enzymes is of special interest as it bears on our understanding of the genesis of glycolysis, one of the oldest and best preserved biochemical pathways. One molecular evolution theory, first proposed by Horowitz<sup>27,28</sup>, postulates that biochemical pathways evolved backwards and that consecutive enzymes that must bind the same substrate/product molecules have evolved by a series of gene duplication events, with the consequence that all the members of a particular pathway are related to one another. Divergent evolution of PK and enolase, two consecutive enzymes in the glycolytic pathway, supports this theory.

We thank Drs R. H. Kretsinger and R. T. Chandross for collecting the data, Dr H. Muirhead for refined coordinates of pyruvate kinase, Dr A. Goldman for coordinates of muconate lactonizing enzyme, and Drs A. Tulinsky, J. M. Bewer and M. H. Hatada for helpful discussions. This work was supported by grants from the NIH (to L.L.).

Received 31 March; accepted 29 April 1988.

- Lebioda, L. & Brewer, J. M. *J. molec. Biol.* **180**, 213-215 (1984).
- Wang, B.-C. *Meth. Enzym.* **115**, 90-111 (1985).
- Chin, C. C. Q., Brewer, J. M. & Wold, F. *J. biol. Chem.* **256**, 1377-1384 (1981).
- Holland, M. J., Holland, J. P., Thill, G. P. & Jackson, K. A. *J. biol. Chem.* **256**, 1385-1395 (1981).
- Brewer, J. M. *Crit. Rev. Biochem.* **11**, 209-254 (1981).
- Elliott, J. I. & Brewer, J. M. *Arch. Biochem. Biophys.* **192**, 203-213 (1979).
- Banner, D. W. *et al. Nature* **255**, 609-614 (1975).
- Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. *J. molec. Biol.* **134**, 109-142 (1979).
- Mavridis, J. M., Hatada, M. H., Tulinsky, A. & Lebioda, L. *J. molec. Biol.* **162**, 419-444 (1982).
- Syngusch, J., Beaudry, D. & Allaire, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7846-7850 (1987).
- Matsuura, Y., Kusunoki, M., Harada, W. & Kakudo, M. *J. Biochem.* **95**, 697-702 (1984).
- Carrell, H. L., Rubin, B. H., Hurley, T. J. & Glusker, J. P. *J. biol. Chem.* **259**, 3230-3236 (1984).
- Lindqvist, Y. & Brändén, C.-I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6855-6859 (1985).
- Schneider, G., Lindqvist, Y., Brändén, C.-I. & Lorimer, G. *EMBO J.* **5**, 3409-3415 (1986).
- Goldman, A., Ollis, D. L. & Steitz, T. A. *J. molec. Biol.* **194**, 143-153 (1987).
- Hyde, C. C., Padlan, E. A., Ahmed, S. A., Miles, E. W. & Davies, D. R. *Fedn Proc.* **46**, 2215 (1987).
- Lim, L. W. *et al. J. biol. Chem.* **261**, 15140-15146 (1986).
- Xia, Z. X. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 2629-2633 (1987).
- Priestle, J. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 5690-5694 (1987).
- Rossmann, M. G. *Phil. Trans. R. Soc. B293*, 191-203 (1981).
- Lebioda, L., Hatada, M. H., Tulinsky, A. & Mavridis, I. M. *J. molec. Biol.* **162**, 445-458 (1982).
- Muirhead, H. *Trends biochem. Sci.* **8**, 326-330 (1983).
- Fothergill-Gilmore, L. A. in *Multidomain Proteins—Structure and Evolution* (ed. Hardie, D. G. & Coggins, J. R.) 85-174 (Elsevier, Amsterdam, 1986).



24. Schulz, G. E. *Angew. Chem.* **20**, 143-151 (1981).
25. Yeh, W. K. & Ornston, L. N. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5365-5369 (1980).
26. Rose, I. A. *Meth. Enzym.* **87**, 84-97 (1982).
27. Horowitz, N. H. *Proc. natn. Acad. Sci. U.S.A.* **31**, 153-156 (1945).
28. Horowitz, N. H. in *Evolving Genes and Proteins* (ed. Bryson, V. & Vogel, H. J.) 15-26 (Academic, New York, 1965).
29. Rossmann, M. G. & Argos, P. *J. biol. Chem.* **250**, 7525-7532 (1975).

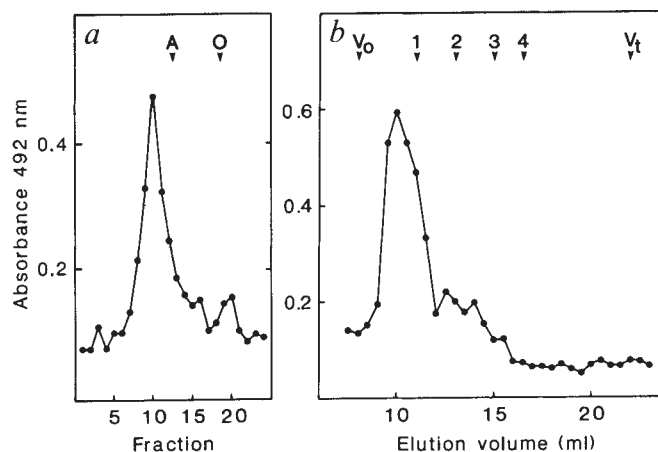
## Requirement of hormone for thermal conversion of the glucocorticoid receptor to a DNA-binding state

Marc Denis, Lorenz Poellinger, Ann-Charlotte Wikstöm & Jan-Åke Gustafsson\*

Department of Medical Nutrition, Karolinska Institute, Huddinge University Hospital F-69, S-14186 Huddinge, Sweden

A central question arising from the model of eukaryotic gene regulation by steroid hormone receptors is whether or not proteins represent pre-existing gene regulatory proteins that are activated on exposure to the extracellular signal. It has been generally believed that the ligand-binding of steroid hormone receptors triggers an allosteric change in receptor structure, manifested by an increased affinity of the receptor for DNA *in vitro* and nuclear target elements *in vivo*, as monitored by nuclear translocation<sup>1,2</sup>. But this model has been challenged by recent reports indicating that glucocorticoid and progesterone receptors bind specifically *in vitro* to target DNA sequences even in the absence of hormone<sup>3,4</sup>. On the other hand, it appears that the hormone induces protection *in vivo* of the glucocorticoid response element of the tyrosine amino transferase gene<sup>5</sup>. Here we show that under conditions permitting minimal *in vitro* manipulation, the steroid-free glucocorticoid receptor in crude cytosol associates with the hsp90 heat shock protein (relative molecular mass  $M_r \approx 90,000$ ) to form a large 300K complex, rather than the 94K liganded receptor monomer. More importantly, we have developed an assay to demonstrate the requirement of hormone to dissociate the 300K complex by heat treatment. Specific DNA-binding activity of the receptor becomes apparent in this process, showing that DNA binding occurs but is inhibited in the large heteromeric complex. We propose a model in which receptor function is repressed by association of the receptor with hsp90. Dissociation of this complex is induced by the binding of steroid and is apparently an irreversible process.

The dissociation of hsp90 is considered necessary for the nonspecific DNA binding of the 94K receptor monomer<sup>6,7</sup>, on the basis that the glucocorticoid receptor is stabilized by sodium molybdate in the presence of ligand in a non-DNA-binding (non-activated) form as a 300K (~9S) complex<sup>8</sup>. In an effort to reconcile the apparently conflicting data on whether or not hormone is required for DNA binding of the receptor, we examined the hydrodynamic and DNA-binding properties of



**Fig. 1** Hydrodynamic characterization of the non-liganded glucocorticoid receptor under molybdate-free conditions. *a*, Samples of untreated crude cytosol were analysed on 5–20% sucrose gradients. Aldolase (A, 7.8 S) and ovalbumin (O, 3.0 S) were used as external standards and run on separate gradients. *b*, Aliquots of crude cytosol were chromatographed on a Superose 12 column calibrated with blue dextran ( $V_0$ ),  $^3\text{H}_2\text{O}$  ( $V_t$ ) and the following proteins: 1,  $\beta$ -galactosidase (6.9 nm); 2, aldolase (4.8 nm); 3, ovalbumin (3.0 nm); and 4, myoglobin (2.0 nm). Fractions were assayed for glucocorticoid receptor content by direct ELISA. A high absorbance at 492 nm corresponds to large amounts of receptor in the fraction.

**Methods.** Eight-week-old male Sprague-Dawley rats were adrenalectomized five days before killing. Livers were perfused and cytosol was prepared in EPG buffer (20 mM sodium phosphate, 1 mM EDTA, 10% glycerol, pH 7.4) containing 50 mM NaCl and 2 mM dithiothreitol<sup>13</sup>. Samples (200  $\mu\text{l}$ ) were loaded on the top of 5–20% sucrose gradients prepared in EPG buffer containing 150 mM NaCl. Centrifugation was performed at 406,000g for 95 min at 2°C in a Beckman Vt80 rotor. Five-drop fractions were collected, and aliquots assayed for glucocorticoid receptor content. Gel permeation chromatography was performed on a Pharmacia Superose 12 HR 10/30 column equilibrated with 20 mM sodium phosphate, 1 mM EDTA, 150 mM NaCl, pH 7.4. The flow rate was 0.5 ml per min, and fractions were collected every min. Aliquots (50  $\mu\text{l}$ ) of fractions obtained after sucrose gradient centrifugation or gel permeation chromatography were incubated for 1 h in 96-well ELISA plates coated with 100 ng of purified monoclonal antibody<sup>14</sup> and blocked with gelatin. The presence of glucocorticoid receptor was then assayed for by the following reagents: (1) rabbit anti-rat glucocorticoid receptor diluted 1:400; (2) swine anti-rabbit immunoglobulins, horseradish peroxidase conjugated, diluted 1:200; and (3) hydrogen peroxide and 1,2-phenylenediamine dihydrochloride. The product of the enzyme reaction was monitored spectrophotometrically at 492 nm.

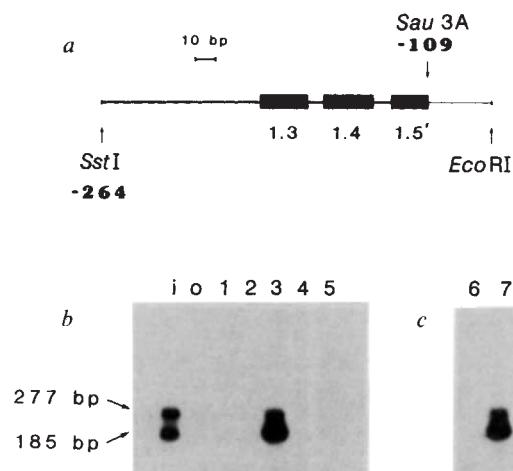
the steroid-free glucocorticoid receptor after a number of different *in vitro* treatments. Following analysis, we assayed for glucocorticoid receptor by an enzyme-linked immunosorbent assay (ELISA) based on antibodies raised against the rat liver glucocorticoid receptor<sup>9</sup>. We detected a peak of receptor immunoreactivity sedimenting at ~9S in sucrose gradients, even under molybdate-free conditions at an ionic strength of 150 mM NaCl (Fig. 1*a*). Analysis of the unliganded receptor by gel permeation high performance liquid chromatography revealed a Stokes' radius of ~8 nm (Fig. 1*b*). The properties of this complex are identical to the molybdate-stabilized heteromeric complex of receptor and hsp90 (compare above)<sup>10</sup>.

Activation of the liganded glucocorticoid receptor has been shown to be temperature-dependent<sup>11</sup>. To assess the DNA-binding activity of the steroid-free receptor, crude cytosol was chromatographed on DNA-cellulose after exposure to 25°C for 30 min. Glucocorticoid receptor was only detectable by ELISA in the flow-through fractions of the column (data not shown)<sup>12</sup>.

\* To whom correspondence should be addressed.

**Table 1** Treatments of the glucocorticoid receptor in crude cytosol

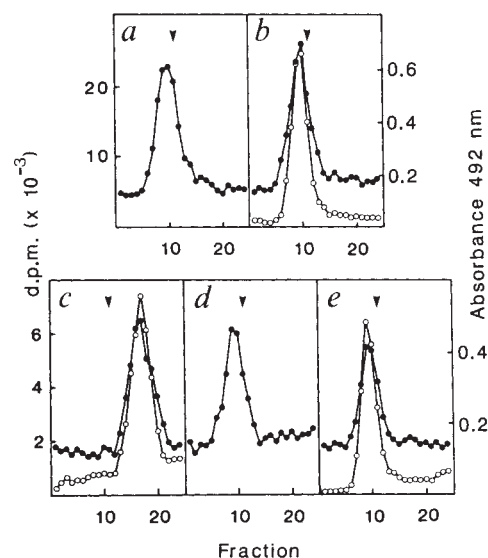
1. No treatment
2. Labelling with [ $^3\text{H}$ ]triamcinolone acetonide (100 nM, 4°C, 1 h)
3. Labelling with [ $^3\text{H}$ ]triamcinolone acetonide (100 nM, 4°C, 1 h), followed by exposure to elevated temperature (25°C, 30 min)
4. Exposure to elevated temperature (25°C, 30 min) in the absence of steroid
5. Exposure to elevated temperature (25°C, 30 min) followed by labelling with [ $^3\text{H}$ ]triamcinolone acetonide (100 nM, 4°C, 1 h)



**Fig. 2** Importance of hormone and temperature for specific DNA-binding properties of the glucocorticoid receptor. Samples of liver cytosol, treated as described in Table 1, were supplemented with 10 mM sodium molybdate and were incubated with an equimolar mixture of two radioactive DNA fragments containing specific glucocorticoid receptor-binding sites (*a*, black bars) from MMTV or nonspecific DNA, respectively. Glucocorticoid receptors were immunoadsorbed on an anti-glucocorticoid receptor monoclonal antibody coupled to CNBr-activated Sepharose 4B (ref. 14). The co-precipitated DNA fragments were analysed on a 5% polyacrylamide gel and visualized by autoradiography (*b*). In lane *i* is shown an aliquot of the input mixture of the two DNA fragments used. For lane 0, EPG buffer was used instead of cytosol. Lanes 1-5 correspond to the different conditions of treatment (see Table 1). *c*, After adsorption of unliganded crude glucocorticoid receptor on the immunosorbent, the gel was washed in buffer with (lane 6) or without (lane 7) molybdate, and incubated with the mixture of the two labelled DNA fragments. The retained DNA fragments were analysed as in *b*.

**Methods.** A 185-bp *SstI*-*EcoRI* fragment was isolated from the plasmid pLS5'139 (ref. 13). This fragment contains three individual glucocorticoid receptor binding sites within the 5' long terminal repeat of MMTV as assessed by DNase I and methylation protection experiments<sup>13,28</sup>. As a nonspecific control fragment, a 277-bp *Bam*HI-*Hinc*II fragment which does not contain any consensus sequence motifs for glucocorticoid receptor binding was isolated from pBR322. The fragments were 3'-labelled using [ $\alpha$ -<sup>32</sup>P]dATP and the Klenow fragment of DNA polymerase I. *b*, Labelled fragments (5 ng) were incubated with cytosol previously exposed to the treatments summarized in Table 1 and subsequently supplemented with 10 mM sodium molybdate to prevent any further activation process. Incubations were carried out in a total volume of 100  $\mu$ l for 30 min at 4°C. Following addition of immunosorbent (30  $\mu$ l), the incubation was continued at 4°C for 45 min. The gels were then pelleted by centrifugation and washed 3 times with 1 ml EPGM<sub>0</sub> buffer (10 mM sodium molybdate in EPG buffer). Co-precipitated DNA was extracted with 0.1% sodium dodecyl sulphate in EPG buffer for 15 min at 65°C, precipitated with ethanol, and analysed by 5% (w/v) polyacrylamide gel electrophoresis and autoradiography. *c*, Aliquots of crude cytosol (100  $\mu$ l) were incubated with 30  $\mu$ l of immunosorbent for 45 min at 4°C. The gels were then washed 3 times with 1 ml EPGM<sub>0</sub> (lane 6) or EPG (lane 7) buffer containing 150 mM NaCl. Incubation with labelled DNA fragments, extraction and analysis of precipitated fragments was performed as in *b*.

Moreover, analysis of these fractions on sucrose gradients showed that the glucocorticoid receptor sedimented at ~9S (data not shown). The effect of exposure to hormone and/or to elevated temperature on the conversion of the glucocorticoid receptor to a DNA-binding form was then investigated under a variety of conditions (summarized in Table 1). For specific DNA-binding experiments, we used a 185-base pair (bp) fragment containing three glucocorticoid receptor-binding sites from the 5' long terminal repeat of mouse mammary tumour virus



**Fig. 3** Hormone- and temperature-dependent changes in sedimentation behaviour of the glucocorticoid receptor. Samples (1 ml) of liver cytosol were submitted to the different treatments summarized in Table 1. Following chromatography on PD-10 columns (Pharmacia) equilibrated with EPGM<sub>0</sub> buffer to remove any unbound steroid, 200  $\mu$ l aliquots were run on 5-20% sucrose gradients prepared in EPGM<sub>0</sub> buffer containing 150 mM NaCl. Centrifugation was as described in Fig. 1. Collected fractions were then analysed for glucocorticoid receptor immunoreactivity by ELISA (●) and 50  $\mu$ l-aliquots were counted (○). The position of aldolase (7.8 S) is indicated by an arrow.

(MMTV)<sup>13</sup>. The cytosol was treated, and samples were incubated with an equimolar concentration of these two fragments end-labelled to a similar specific activity. Receptor-DNA complexes were immunoprecipitated by an anti-glucocorticoid receptor monoclonal antibody coupled to Sepharose<sup>14</sup>, and the precipitated DNA was analysed by polyacrylamide gel electrophoresis. The monoclonal antibody we used (number 7 in ref. 9) recognizes the glucocorticoid receptor in its DNA-binding<sup>9</sup>, molybdate-stabilized<sup>14</sup> and unliganded<sup>15</sup> states. Figure 2 clearly shows that interaction of the glucocorticoid receptor with MMTV DNA only occurred in the presence of hormone and after exposure to elevated temperature. In agreement with its behaviour during DNA-cellulose chromatography, the heat-treated unliganded receptor did not bind to specific DNA, even when the receptor was labelled following heat treatment. Finally, independently of ligand, preincubation of cytosol with sodium molybdate gave negative DNA-binding results.

A complete correlation was observed between the shift in sedimentation position from ~9S to ~4S of the glucocorticoid receptor and its acquisition of a specific DNA-binding state. Thus, the glucocorticoid receptor sedimented at ~4 S only after preincubation with hormone followed by heat-treatment (Fig. 3). This observation was confirmed by analysis of the glucocorticoid receptor by high-performance anion-exchange chromatography<sup>16</sup> after an identical pattern of treatments. In these experiments, the glucocorticoid receptor represented a more basic species only under conditions resulting in DNA binding, and a sedimentation position of ~4 S (data not shown).

In this report, we show for the first time that hsp90 represses receptor function in a hormone-dependent manner as assessed by specific DNA binding. Recently it has been proposed that binding of hormone is a prerequisite for temperature-induced dissociation of hsp90 (ref. 17) and salt-induced sedimentation-shift of the receptor<sup>18</sup>. The treated cytosol was incubated with



hormone and analysed by sucrose gradient centrifugation<sup>18</sup>, but the receptor was not analysed for its functional property of specific DNA binding. Moreover, in studies relying on steroid binding for receptor detection, the possibility that the receptor has lost its hormone-binding capacity<sup>19</sup> during treatment cannot be excluded. In the results we present here, glucocorticoid receptor was detected independently of steroid binding using immunochemical techniques.

Artificial dissociation of subunits may account for the previously reported DNA-binding activity of the steroid-free glucocorticoid receptor<sup>3</sup> and the immunoaffinity purified<sup>20</sup> progesterone receptor<sup>4</sup>. In support of this idea, extensive washing with molybdate-free salt-containing buffer dissociates the hsp90 from the immunoadsorbed steroid-free 94K receptor monomer, which then binds specifically to DNA (Fig. 2c). In fact, the 300K, non-DNA-binding form of the unoccupied receptor is labile *in vitro*, and dissociation of the heteromeric receptor to give DNA-binding of the steroid-free receptor monomer will occur during prolonged operations such as gradient centrifugation for 16 h (data not shown), gel permeation chromatography<sup>12</sup> and ammonium sulphate precipitation<sup>17</sup>.

We have tested the reversibility of the activation of glucocorticoid-receptor complexes to a DNA-binding state by the dissociation of hsp90. The liganded DNA-binding form of the glucocorticoid receptor was purified by DNA-cellulose chromatography<sup>13</sup>, and was found to sediment at ~4S, even after incubation with purified hsp90 (ref. 14) in the presence of molybdate (data not shown). Furthermore, when activated DNA-binding glucocorticoid receptor was adsorbed on the immunoaffinity matrix (see above) and incubated with a molar excess of a homogeneous population of hsp90, the DNA-binding activity of the glucocorticoid receptor was not decreased (data not shown). These results indicate that the dissociation of the 9 S non-DNA-binding heteromer to a 4 S DNA-binding species is probably an irreversible process. Interestingly, we (M.D., A.-C.W. and J.Å.G., manuscript in preparation) and others<sup>21</sup> have found that hsp90 interacts with the steroid-binding domain of the glucocorticoid receptor. Thus, hormone-binding may induce a conformational change within the steroid-binding domain of the receptor which prevents any further reassociation of the hsp90 with the glucocorticoid receptor following activation. In support of this contention, expression of glucocorticoid receptor fragments lacking the steroid-binding domain yields proteins which are constitutive with regard to DNA-binding activity and transcriptional activation<sup>22,23</sup>. Finally, steroid-binding has also been proposed to result in an irreversible activation of the progesterone receptor<sup>24</sup>.

In conclusion, our results strongly indicate that the glucocorticoid receptor can be converted to a form interacting with target DNA solely in its liganded state. Moreover, the data support a model where hormone induction of DNA-binding, by modification of a distinct precursor form, precedes activation of target gene expression. Possibly this modification involves unmasking of the previously obscured DNA-binding site on the receptor protein by dissociation of a larger heteromeric complex. Recently it has been shown that induction of sequence-specific binding of both the heat shock transcription factor<sup>25,26</sup> and the kappa-immunoglobulin enhancer-binding factor<sup>27</sup> occurs in the absence of protein synthesis. Therefore, post-translational modification of pre-existing regulatory factors to DNA-binding forms may represent a general mechanism in regulation of inducible gene expression.

*In vivo*, protection against methylation of the glucocorticoid response elements appears to be steroid-dependent. These *in vivo* observations strongly support the significance of the *in vitro* properties of the receptor described here. Clearly, a detailed *in vitro* dissection of the molecular events leading to functional forms of the glucocorticoid receptor is necessary for further elucidation of the mechanism of gene regulation by steroid hormones.

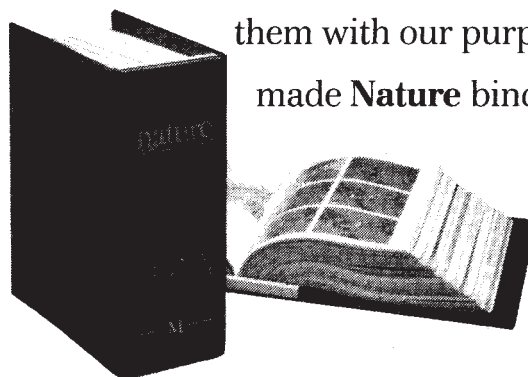
This work was supported by the Swedish Medical Research Council. L.P. and A.-C.W. are recipients of research fellowships from the Swedish Medical Research Council.

Received 10 November 1987; accepted 11 May 1988.

1. Yamamoto, K. R. & Alberts, B. M. *A. Rev. Biochem.* **45**, 721-746 (1976).
2. Gustafsson, J.-Å. *et al. Endocrine Rev.* **8**, 185-234 (1987).
3. Willmann, T. & Beato, M. *Nature* **324**, 688-691 (1986).
4. Bailly, A., Le Page, C., Rauch, M. & Milgrom, E. *EMBO J.* **5**, 3235-3241 (1986).
5. Becker, P. B., Gloss, B., Schmid, W., Strähle, U. & Schütz, G. *Nature* **324**, 686-688 (1986).
6. Housley, P. R., Sanchez, E. R., Westphal, H. M., Beato, M. & Pratt, W. B. *J. biol. Chem.* **260**, 13810-13817 (1985).
7. Mendel, D. B., Bodwell, J. E., Gametchu, B., Harrison, R. W. & Munck, A. *J. biol. Chem.* **261**, 3758-3763 (1986).
8. Dahmer, M. K., Housley, P. R. & Pratt, W. B. *A. Rev. Physiol.* **46**, 67-81 (1984).
9. Okret, S., Wikström, A.-C., Wrangé, Ö., Andersson, B. & Gustafsson, J.-Å. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1609-1613 (1984).
10. Wilhelmsson, A., Wikström, A.-C. & Poellinger, L. *J. biol. Chem.* **261**, 13456-13463 (1986).
11. Schmidt, T. J. & Litwack, G. *Physiol. Rev.* **62**, 1131-1192 (1982).
12. Radojicic, M., Okret, S., Wrangé, Ö. & Gustafsson, J.-Å. *J. Steroid Biochem.* **23**, 1-8 (1985).
13. Wrangé, Ö., Carlstedt-Duke, J. & Gustafsson, J.-Å. *J. biol. Chem.* **261**, 11770-11778 (1986).
14. Denis, M., Wikström, A.-C. & Gustafsson, J.-Å. *J. biol. Chem.* **262**, 11803-11806 (1987).
15. Wikström, A.-C., Bakke, O., Okret, S., Brönnegård, M. & Gustafsson, J.-Å. *Endocrinology* **120**, 1232-1242 (1987).
16. Denis, M., Guendouz, F., Blanchardie, P., Bernard, S. & Lustenberger, P. *J. Steroid Biochem.* **30**, 307-312 (1988).
17. Sanchez, E. R. *et al. J. biol. Chem.* **262**, 6986-6991 (1987).
18. Groyer, A., Schweizer-Grover, G., Cadepond, F., Mariller, M. & Baulieu, E. E. *Nature* **328**, 624-626 (1987).
19. Mendel, D. B., Bodwell, J. E. & Munck, A. *Nature* **324**, 478-480 (1986).
20. Logeat, F. *et al. Biochemistry* **24**, 1029-1035 (1985).
21. Pratt, W. B. *et al. J. biol. Chem.* **263**, 267-273 (1988).
22. Godowski, P. J., Rusconi, S., Miesfeld, R. & Yamamoto, K. R. *Nature* **325**, 365-368 (1987).
23. Rusconi, S. & Yamamoto, K. R. *EMBO J.* **6**, 1309-1315 (1987).
24. Renoir, J.-M., Buchou, T. & Baulieu, E.-E. *Biochemistry* **25**, 6405-6413 (1986).
25. Zimarino, V. & Wu, C. *Nature* **327**, 727-730 (1987).
26. Kingston, R. E., Schuetz, T. J. & Larin, Z. *Molec. cell. Biol.* **7**, 1530-1534 (1987).
27. Sen, R. & Baltimore, D. *Cell* **47**, 921-928 (1987).
28. Payvar, F. *et al. Cell* **35**, 381-392 (1983).

## Give your copies of *Nature* the treatment they deserve

Organise and protect them with our purpose made **Nature** binders



Single binder: UK £5.25; Rest of World £7.75  
Set of four (takes one year's issues): UK £16.00;  
Rest of World £22.00 (prices include postage and packing)

Orders with payment to:  
**Nature**, Circulation Department  
Brunel Road, Basingstoke  
Hants RG21 2XS, UK.  
(Cheques should be made payable to **Nature**)